Marzano Research

DECEMBER 2019

IMPROVING STUDENT LEARNING AND ENGAGEMENT THROUGH GAMIFIED INSTRUCTION: EVALUATION OF IPERSONALIZE

PREPARED BY

MARY KLUTE, DAVID YANOSKI,
CHRISTOPHER RHOADS, JENNIFER NORFORD,
JEANETTE JOYCE, AND MARINA SERDIOUK

Referencing This Report

This report should be cited as: Klute, M., Yanoski, D., Rhoads, C., Norford, J., Joyce, J., & Serdiouk, M. (2019). *Improving student learning and engagement through gamified instruction: Evaluation of iPersonalize*. Centennial, CO: Marzano Research.

Acknowledgements

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305L170006 to Marzano Research. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. We would like to thank the staff, teachers, and students at Fullerton School District for their contributions to this research. We also thank Charles Harding, Lyn Bopp, and Kyle Luke for their contributions to this report. We are also grateful to Helen Apthorp and Richard Mayer for their comments on an earlier version of this report.

Improving Student Learning and Engagement Through Gamified Instruction

Evaluation of iPersonalize

Prepared by

Mary Klute
David Yanoski
Christopher Rhoads
Jennifer Norford
Jeanette Joyce
Marina Serdiouk

CONTENTS

Executive Summary 1
Introduction 4
iPersonalize 4
iPersonalize Instructional Cycle 5
The Study 7
Methods 9
Participants 9
Measures 10
i-Ready Diagnostic 10
MY Access! School Edition 11
YouthTruth Feedback for Teachers Survey 11
iPersonalize Usage Data 12
Implementation of iPersonalize During the Study 12
Results 16
Research Question 1—What Is the Impact of iPersonalize on Average Sixth-Grade Student Achievement in ELA? 16
Research Question 2—What Is the Impact of iPersonalize on Average Sixth-Grade Student Engagement in School? 17
Research Question 3—To What Extent Does Gender Moderate the Impact of iPersonalize on Student Achievement or Student Engagement? 18
Research Question 4—Is the Extent of Use of iPersonalize Associated with Increased Student Achievement? 18
Research Question 5—What Is the Impact of iPersonalize on Low and High Quantiles of Sixth-Grade Student Achievement in ELA and Student Engagement in School? 20
Conclusions 21
Lessons Learned 21
Appendix A. Study Design and Research Methods 23
Random Assignment of Teachers 23
Data Analysis Procedures 24
Baseline Equivalence 25

Analyses to Address Research Questions 25
Appendix B. Descriptive Statistics for Study Measures 30
Appendix C. Detailed Results for Statistical Models 31
Research Question 1—What Is the Impact of iPersonalize on Average Sixth-Grade Student Achievement in ELA? 31
Research Question 2—What Is the Impact of iPersonalize on Average Sixth-Grade Student Engagement in School? 34
Research Question 3—To What Extent Does Gender Moderate the Impact of iPersonalize on Student Achievement or Student Engagement? 37
Research Question 4—Is the Extent of Use of iPersonalize Associated with Increased Student Achievement? 44
Research Question 5—What Is the Impact of iPersonalize on Low and High Quantiles of Sixth-Grade Student Achievement in ELA? 56
References 58

TABLES AND FIGURES

Figure 1. iPersonalize Instructional Cycle ... 7

Table 1. Sample Characteristics by Condition ... 9

Table 2. Alignment of iPersonalize to Computer Game Elements (Landers, 2015) ... 13

Figure 2. Average Change in Reading Scale Scores from Pretest to Post-Test, by Classroom Type ... 17

Table 3. Summary of Results of Models Examining the Association Between Student Usage Variables and Academic Outcomes ... 19

Table A1. Number of Classrooms, by Condition ... 24

Table B1. Descriptive Statistics for Academic and Engagement Outcomes ... 30

Table B2. Descriptive Statistics for iPersonalize Usage Variables (treatment group only) ... 30

Table C1. HLM Coefficients for Model Predicting Reading Scale Score (*n* = 42 classrooms, 1,247 students) ... 31

Table C2. HLM Coefficients for Model Predicting Writing Holistic Score (*n* = 42 classrooms, 1,151 students) ... 33

Table C3. HLM Coefficients for Model Predicting Student Engagement (*n* = 40 classrooms, 1,137 students) ... 34

Table C4. WLS Regression Coefficients for Model Predicting Classroom-Level Engagement (*n* = 42 classrooms) ... 36

Table C5. HLM Coefficients for Model Predicting Reading Scale Score with Gender as a Moderating Variable (*n* = 42 classrooms, 1,247 students) ... 37

Table C6. HLM Coefficients for Model Predicting Writing Holistic Score with Gender as a Moderating Variable (*n* = 42 classrooms, 1,151 students) ... 39

Table C7. HLM Coefficients for Model Predicting YouthTruth Score with Gender as a Moderating Variable (*n* = 40 classrooms, 1,137 students) ... 40

Table C8. HLM Coefficients for Model Predicting YouthTruth Score, Boys Only (*n* = 39 classrooms, 557 students) ... 42

Table C9. HLM Coefficients for Model Predicting YouthTruth Score, Girls Only (*n* = 40 classrooms, 580 students) ... 43

Table C10. HLM Coefficients for Model Examining the Association Between the Number of Log-ons and Reading Scale Score (*n* = 21 classrooms, 601 students) ... 45

Table C11. HLM Coefficients for Model Examining the Association Between the Number of Attempted Quests and Reading Scale Score (*n* = 21 classrooms, 601 students) ... 46

Table C12. HLM Coefficients for Model Examining the Association Between the Number of Completed Quests and Reading Scale Score (n = 21 classrooms, 601 students)............................ 47

Table C13. HLM Coefficients for Model Examining the Association Between the Percentage of Attempted Quests That Were Completed and Reading Scale Score (n = 21 classrooms, 601 students)... 48

Table C14. HLM Coefficients for Model Examining the Association Between the Number of Logons and Writing Holistic Score (n = 21 classrooms, 569 students)... 50

Table C15. HLM Coefficients for Model Examining the Association Between the Number of Attempted Quests and Writing Holistic Score (n = 21 classrooms, 569 students).......................... 51

Table C16. HLM Coefficients for Model Examining the Association Between the Number of Completed Quests and Writing Holistic Score (n = 21 classrooms, 569 students).......................... 52

Table C17. HLM Coefficients for Model Examining the Association Between the Percentage of Attempted Quests That Were Completed and Writing Holistic Score (n = 21 classrooms, 569 students)... 53

Table C18. Summary of Results of Models Examining the Causal Relationship Between Student Usage Variables and Academic Outcomes 55

Table C19. Summary of Quantile Mixed Models Examining Impact of Randomization to iPersonalize on Academic Outcomes at Percentiles of Conditional Achievement Distribution... 56

Table C20. Summary of Quantile Mixed Models Examining Impact of Randomization to iPersonalize on Engagement at Percentiles of Conditional Engagement Distribution. 57

EXECUTIVE SUMMARY

The purpose of this research study was to evaluate iPersonalize, a gamified instructional approach developed by Fullerton School District (FSD) in California to encourage student engagement and promote achievement. An instructional approach is considered gamified when it incorporates computer game elements to augment existing classroom, instructional, and assessment processes (Bedwell, Pavlas, Heyne, Lazzara, & Salas, 2012; Landers, 2015). In developing iPersonalize, FSD reimagined the organization, learning activities, instruction, and assessments of a typical unit of study by employing the principles of alternate reality games (ARGs). ARGs weave together real-world and online experiences through a narrative, and participants complete tasks or quests related to that narrative both in the real world and online (Educause, 2009). Using ARG principles, FSD developed instructional units around an "epic storyline," a story or scenario that anchors instruction and binds all instruction and assessment activities together. FSD expected these features to lead to increased student engagement in learning, which in turn would support student achievement.

The study employed a randomized controlled trial designed to support causal inferences about the effectiveness of iPersonalize for impacting sixth-grade student engagement and achievement in English language arts (ELA). ELA achievement is of practical interest to schools, districts, and states as they strive to prepare students with effective written and spoken communication skills for participation in the global economy. Student engagement is of practical interest to educators because it is a strong predictor of student achievement (Appleton, Christenson, Kim, & Reschly, 2006), and engagement tends to decline as students get older (Fredricks, Blumenfeld, & Paris, 2004; Fredricks & McColskey, 2012). Both engagement and achievement were identified as targets of iPersonalize. This report summarizes the results of the study, which was conducted in fall 2017.

The study focused on two main research questions:

1. What is the impact of iPersonalize on average sixth-grade student achievement in ELA?
2. What is the impact of iPersonalize on average sixth-grade student engagement in school?

The study also addressed three exploratory research questions:

3. To what extent does gender moderate the impact of iPersonalize on student achievement or student engagement?
4. Is the extent of use of iPersonalize associated with increased student achievement?
5. What is the impact of iPersonalize at low and high quantiles of sixth-grade student achievement in ELA and student engagement in school?

The study included 1,295 students from 42 classrooms in 15 schools. All students were enrolled in sixth grade in FSD during the 2017/18 school year. Students in 24 of these classrooms were assigned to ELA instruction using iPersonalize. Students in the remaining 18 classrooms were assigned to business-as-usual instruction. Teachers in both groups were expected to teach the same ELA unit. Teachers in the iPersonalize group were expected to incorporate elements of gamification, while the teachers in the control group were expected to not incorporate elements of gamification.

There were 721 students in the treatment group and 574 students in the control group. The study used district-administered assessments of reading, writing, and student engagement. For students assigned to the iPersonalize condition, usage data from the iPersonalize online learning management system were also used for the study.

Key findings from the study were as follows:

- On both reading and writing assessments, the difference between the treatment group and the control group was small and not statistically significant, indicating that the two groups performed similarly.
- Students in both groups reported similar levels of engagement in school.
- Gender did not significantly moderate the impact of iPersonalize on student achievement or student engagement.
- The impact of the program on reading and writing assessments was close to zero, regardless of the extent to which students interacted with the online learning management system.
- There was some evidence to suggest that the program had a stronger impact on engagement for students who were already the most engaged in school.

During fall 2017, FSD implemented a newly developed iPersonalize unit that centered on passion projects. This unit deviated from previous implementations of iPersonalize in two ways that are relevant to interpreting the results of the study:

1. The passion projects were not directly linked to the ELA standards. In effect, two units were running in parallel: one focused on helping students identify and explore their interests and strengths, and one focused on ELA standards such as word choice, supporting evidence, and pronouns.
2. The passion projects were not carried out in the context of an epic storyline or quest.

Unfortunately, issues with implementation of iPersonalize hampered the ability of the study to detect effects of the intervention. Due to demands from teachers, both the treatment and control groups implemented most of the features of the iPersonalize approach. Although the students assigned to the control group did not have access to the iPersonalize online learning management system, they did participate in other online activities and received instruction that incorporated game elements. Because the

differences in treatment between the two groups were negligible, it is not surprising that there were no discernable differences in academic achievement and engagement between students assigned to the iPersonalize group and those assigned to business-as-usual instruction. However, students in both groups made gains on the reading and writing assessments.

INTRODUCTION

The purpose of this research study was to evaluate iPersonalize, a gamified instructional approach developed by FSD to encourage student engagement and promote achievement. The study employed a randomized controlled trial designed to examine the impact of iPersonalize on sixth-grade student engagement and achievement in ELA. ELA achievement is of practical interest to schools, districts, and states as they strive to prepare students with effective written and spoken communication skills for participation in the global economy. Student engagement is of practical interest to educators because it is a strong predictor of student achievement (Appleton et al., 2006), and engagement tends to decline as students get older (Fredricks et al., 2004; Fredricks & McColskey, 2012). Both engagement and achievement were identified as targets of iPersonalize. This report summarizes the results of the study, which was conducted in fall 2017.

iPersonalize

In 2014, FSD created iPersonalize, an instructional approach designed to promote student engagement and achievement by augmenting current teaching and assessment strategies for ELA and the California Content Standards with the game elements (Box 1). FSD reimagined the organization, learning activities, instruction, and assessments of a typical unit of study by employing game elements. Instruction with iPersonalize was designed to weave together real-world and online experiences through a narrative, with participants completing tasks or quests related to that storyline both in the real world and online. FSD developed instructional units around an “epic storyline,” a story or scenario that anchors instruction and binds all instruction and assessment activities together.

For example, in an iPersonalize writing unit, the storyline focuses on how American forefathers used persuasive writing to inspire colonists to form an army and defeat England in the Revolutionary War. This storyline is intended to help students understand the importance of becoming a skilled persuasive writer before they begin to learn the mechanics of persuasive writing. FSD anticipated that this gamified instructional approach would improve student outcomes within a short period of time (less than one year) because it includes computer game elements that appeal to students and that provide opportunities for frequent practice and feedback. FSD expected these features to lead to increased student engagement in learning, which in turn would support student achievement.

Box 1. Gamification

Gamification, or the use of video game elements in nongaming contexts, is becoming increasingly popular in a variety of contexts such as service marketing, workplace employee development, and K–12 and postsecondary education. Business leaders and educators have used gamification to increase engagement and motivate targeted behaviors. To capitalize on the potential benefits of gamification, instructional designers at both the K–12 and higher education levels are creating instructional activities, assessment and feedback systems, and even entire courses based on game elements. The goal of gamification in education is to increase learner engagement, resulting in improved academic outcomes (Landers, 2015).

An instructional approach is considered *gamified* when it incorporates computer game elements into existing classroom, instructional, and assessment processes. Game elements that can be incorporated include the following:

- Action language, which refers to the ways in which the user interacts with the game.
- Assessment methods that use game points (e.g., experience points or XP) and the tracking of game progress.
- Conflict or challenges that require competition and problem-solving.
- A degree of participant control over the game environment, progress, or choice of activities.
- A change to the learning environment, such as learning in virtual environments or learning outside of the classroom.
- An epic storyline that creates a fictional world in which instructional and assessment activities occur.
- Student interaction in the fictional environment and in collaboration around problem-solving.
- Immersion in the game environment, such as classroom arrangement, decoration, or costumes tied to the epic storyline.
- Rules and goals that are clearly defined and incorporated into classroom procedures. (Bedwell et al., 2012; Landers, 2015)

iPersonalize Instructional Cycle

The iPersonalize instructional cycle includes a teach cycle and a reteach cycle for students who need it. The instructional cycle includes six phases:

1. The *Learning Goals* phase introduces students to the learning goals and standards they are required to meet in the unit. A pre-assessment is also administered.
2. In the *Engage* phase, students are introduced to the epic storyline.

3. The *Interacting with New Knowledge* phase introduces the new content and promotes student interaction through a variety of whole-group, small-group, and individual activities—called *raids* and *quests*—that take place in person, both inside and outside of the classroom, as well as online.
4. A *Common Formative Assessment* (CFA) is delivered to determine student progress.
5. Depending on their CFA results, students take one of two paths:

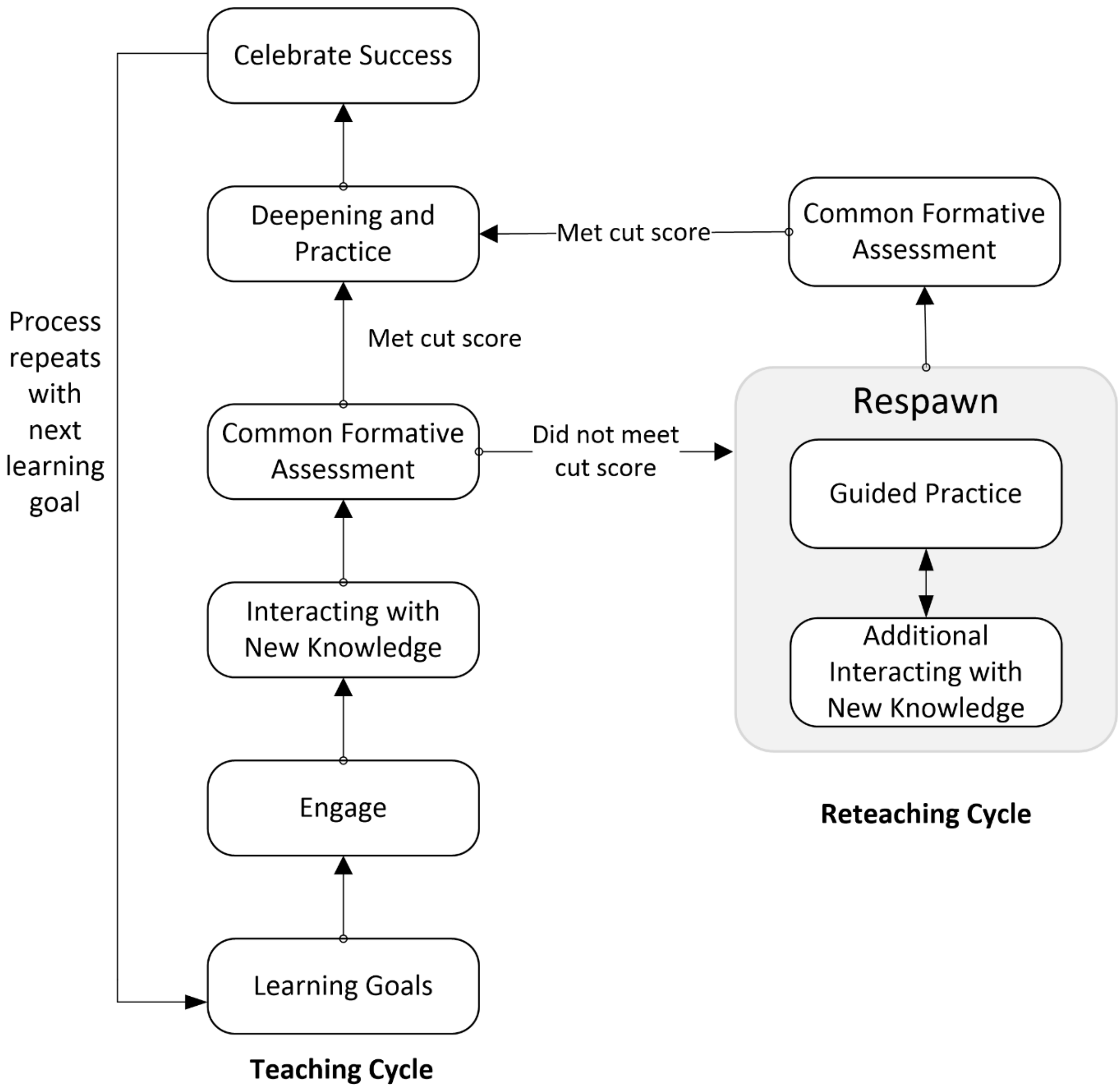
    a. Students who meet the cut score on the CFA unlock and move to the *Deepening and Practicing* phase, in which they complete additional quests, including a secret mission to earn a Met or Mastery badge. A Met badge indicates that they have achieved proficiency on the learning goal according to the learning progression. A Mastery badge indicates that they have applied the newly acquired knowledge in a new or unique situation beyond the requirement for proficiency.
    b. Students who do not meet the cut score are *Respawned* and begin a reteach cycle consisting of additional *Interacting with New Knowledge* quests and raids and *Guided Practice* to work on the content they are struggling with. This reteach cycle continues until students meet the cut score on the CFA and move on to the *Deepening and Practice* phase as described previously.

6. Once students successfully meet the learning goals, they *Celebrate Success* and move on to the next learning goal.

During a typical unit, students may engage in this cycle three to five times with different learning goals (Figure 1).

Implementation of iPersonalize is facilitated by an online learning management system that provides individualized instruction and monitors progress during the *Interacting with New Knowledge*, *Respawn*, and *Deepening and Practicing* phases of the instructional cycle. Using the learning management system, students participate in individual quests aligned to the epic storyline. The learning management system gives feedback in the form of badges for completion of quests, displays student tracking of progress (e.g., XP), and provides extension activities. Students move through quests at their own pace and demonstrate mastery of a standard when they are ready. The iPersonalize approach is expected to foster student engagement and achievement because it offers students some control over their learning pathways while still ensuring that they learn the content of required standards.

Figure 1. iPersonalize Instructional Cycle



THE STUDY

This study used a randomized controlled trial. The goal of the study was to obtain rigorous evidence of the impact of iPersonalize that FSD leadership could use to inform decisions related to scaling up or revising the intervention. Because iPersonalize involves instruction delivered at the classroom level, the unit of randomization was the teacher.

Often, studies that compare gamified and business-as-usual instruction use different approaches to instruction in the two conditions. Such studies have been criticized for confounding the media (i.e., the game) and the instructional methods being used (Mayer, 2014). This study was designed to address such criticism in three ways. First, teachers in both groups were expected to teach the same standardized ELA units. Second, teachers in

both conditions were expected to use the same pretest benchmark assessments on which they could base instructional decisions. Third, iPersonalize was modeled on the research-based framework for effective instruction outlined in *The Art and Science of Teaching* (Marzano, 2007). All FSD teachers, including those randomly assigned to the control group, received professional development on, and were expected to implement the framework described in, *The Art and Science of Teaching*. Given these expectations for implementation, the design of the study would help to isolate the effect of iPersonalize rather than confounding it with the use of particular instructional strategies.

The study focused on two main research questions:

1. What is the impact of iPersonalize on average sixth-grade student achievement in ELA?
2. What is the impact of iPersonalize on average sixth-grade student engagement in school?

The study also addressed three exploratory research questions:

3. To what extent does gender moderate the impact of iPersonalize on student achievement or student engagement?
4. Is the extent of use of iPersonalize associated with increased student achievement?
5. What is the impact of iPersonalize at low and high quantiles of sixth-grade student achievement in ELA and student engagement in school?

Research question 3 was of interest for two reasons. First, girls tend to score higher than boys on reading and writing assessments (Fearrington et al., 2014; National Center for Education Statistics, 2013). Consistent with this research, FSD has also found differences in assessment scores by gender. For example, 61 percent of sixth-grade girls in FSD met or exceeded standards on the 2015 Smarter Balanced ELA assessment, compared with only 51 percent of sixth-grade boys in the district. Second, while data suggest that nearly all boys and girls play computer or video games, boys are about twice as likely to be daily gamers (Lenhart et al., 2008). Given these trends, it is worthwhile to explore whether the use of gamified instruction may be more effective for boys and thus contribute to closing the gender gap in ELA.

Research question 4 was a preliminary attempt at examining, with existing data, the extent to which student engagement with the iPersonalize online learning management system was related to student outcomes. Research question 5 allowed an exploration of whether the impact of iPersonalize differed for students at different percentiles in the achievement and engagement distributions.

METHODS

This section provides an overview of the methods used to conduct the study. Additional technical details about methodology are presented in Appendix A.

PARTICIPANTS

The study included 1,295 students from 42 classrooms in 15 schools. Classroom size ranged from 15 to 35 for the treatment group and 24 to 35 for the control group. All students were enrolled in sixth grade in FSD during the 2017/18 school year. Students in 24 of these classrooms were assigned to ELA instruction using iPersonalize. Students in the remaining 18 classrooms were assigned to business-as-usual instruction (see Appendix A for detailed information about the assignment of classrooms to condition). There were 721 students in the treatment group and 574 students in the control group.

Table 1 presents the characteristics of the overall student sample. Just over half of the students were girls while nearly half qualified for free or reduced-price lunch. About a quarter of students were identified as English learners, 17 percent as gifted students, and 11 percent as special education students. About half of the students identified as Hispanic, and about a third identified as Asian or Pacific Islander. Demographic characteristics were similar for the treatment and control groups.

Table 1. Sample Characteristics by Condition

Sample Characteristic	Overall Sample		Treatment Group		Control Group	
	Number	Percent	Number	Percent	Number	Percent
Female	654	51%	370	51%	284	49%
English learners	310	24%	161	23%	149	26%
Eligible for gifted program	222	17%	144	20%	78	14%
Eligible for special education services	147	11%	79	11%	68	12%
Eligible for free or reduced-price lunch	629	49%	330	46%	299	52%
Race/Ethnicity						
Hispanic	668	52%	342	48%	326	57%
Asian/Pacific Islander	398	31%	257	36%	141	25%
White	179	14%	94	13%	85	15%
Other	45	3%	27	4%	18	3%

Measures

The study focused on two outcome domains: ELA achievement and student engagement. In this study, student scores on two districtwide benchmark assessments were used as the outcome variables for ELA achievement, and one district-administered survey was the outcome for student engagement.

At the beginning of the school year and at the end of the first trimester, FSD administered the i-Ready Diagnostic to monitor student mastery of reading standards and the MY Access! School Edition to monitor student mastery of writing standards. These measures were ideal for the study because they assessed the same reading and writing standards that teachers covered during the first trimester. With such tight alignment, it was reasonable to assume that the measures would be sensitive enough to detect increases in student learning during the first trimester. Pre- and post-test reading and writing scores from the i-Ready Diagnostic and the MY Access! School Edition were used in data analyses.

In contrast to the academic assessments, the YouthTruth Feedback for Teachers Survey was only administered one time during the study.

In addition, to address research question 4, iPersonalize usage data were gathered for the treatment group only. Each of these sources of data are described in more detail below.

i-Ready Diagnostic

i-Ready Diagnostic is a vertically scaled, computer adaptive assessment system. According to the publisher, the assessment

- is aligned to more than 90 percent of the assessable Common Core State Standards in third through eighth grades;
- adheres to *Standards of Psychological and Educational Testing* (American Educational Research Association, American Psychological Association, & National Council of Measurement in Education, 1999);
- was field-tested with over 2 million students;
- has low standard errors of measurement;
- has items that discriminate among students of different abilities; and
- has been approved for use in many areas, including New York, Virginia, Chicago, and Dallas. (Curriculum Associates, n.d.)

Previous research has suggested that i-Ready Diagnostic is strongly predictive of performance on state assessments in ELA. In a correlational study, student performance on the i-Ready Diagnostic was positively and significantly related to performance on the Common Core-aligned New York State ELA test. Correlations of 0.83 and 0.84 were observed for two consecutive cohorts of students (Educational Research Institute of America, 2014).

The i-Ready Diagnostic yields an overall reading scale score along with scale scores for six subscales: high-frequency words; phonics; phonological awareness; reading comprehension: informational text; reading comprehension: literature; and vocabulary. Students in sixth grade are expected to gain between 15 and 23 points on the overall score over the course of the school year (Curriculum Associates, 2017). In this study, the overall reading scale score was used as the primary reading outcome.

MY Access! School Edition

MY Access! is a cloud-based writing instructional and assessment tool that includes over 1,500 prepackaged writing prompts aligned to the Common Core standards. It uses automated scoring based on artificial intelligence and linguistic technologies (IntelliMetric) to score written essays and responses to short-answer questions. Holistic and analytic scores are provided as immediate feedback for students and teachers. For essays by students in sixth through eighth grades, IntelliMetric scores agree with expert rater scores 77 percent of the time, on average, with agreement within one point of nearly 100 percent (Vantage Learning, 2006). For responses to 10 different composition prompts—including *Informational*, *Literary Narrative*, and *Persuasive*—Pearson correlation coefficients between IntelliMetric and expert rater scores ranged from 0.86 to 0.96 (Vantage Learning, 2006). This assessment yields a holistic rating of the quality of writing along with scores for five areas: content and development; language use, voice, and style; mechanics and conventions; organization; and focus and meaning. All scores range from 1 to 6. In this study, the holistic score was used as the primary writing outcome.

YouthTruth Feedback for Teachers Survey

YouthTruth is an online survey developed for students in sixth through twelfth grade. It gathers information about six areas: student engagement; academic rigor and expectations; relevance; instructional methods; personal relationships; and classroom culture. These areas were identified based on early research from the Measures of Effective Teaching study, which indicated that these constructs were highly correlated with effective teaching. The survey was developed with over 2,000 students in 111 classrooms. It has been administered to over 35,000 students in approximately 3,700 classrooms. The measure has good reliability, with alphas ranging from 0.84 to 0.91 (YouthTruth, n.d.).

This study used a shortened version of the YouthTruth survey. It consisted of 17 items rated on a 3-point response scale (1 = no, hardly ever; 2 = sometimes; 3 = yes, very much). Example items included “Do you like coming to your class?”; “Does the work you do in this class make you really think?”; and “Do you learn interesting things in class?” FSD administered the survey in January 2018.[1] All items were summed to form a scale. The

[1] Initially, the plan was to use data from a districtwide administration of the YouthTruth survey at the end of the first trimester. When YouthTruth provided data to FSD in December 2017, it was discovered that

internal consistency of the scale was good (Cronbach's alpha = 0.78). Data for two classrooms could not be linked to individual students.[2]

iPersonalize Usage Data

iPersonalize usage data were available for students assigned to the treatment group only. These data included the number of times each student logged onto the system, the number of quests attempted, the number of quests completed, and the percentage of attempted quests that were completed.

Implementation of iPersonalize During the Study

Rather than implementing a previously developed iPersonalize unit for the study, FSD opted to implement a new unit that was developed during summer 2017. This unit differed in key ways from previous implementations of iPersonalize. This section describes how the unit was implemented, including challenges faced in its implementation.

In fall 2017, FSD implemented a newly developed iPersonalize unit that centered on passion projects. FSD introduced the project in the following announcement on back-to-school night:

> Fullerton School District is proud to announce "Agents of Change," an innovative personalized learning experience. Through this experience, students will discover their interests and strengths while learning California State Standards in Reading, Language and Informative Writing.
>
> Learners in the sixth grade will attend a kickoff event where they will learn more about themselves via a strengths and interests inventory tool called Thrively. Each learner will identify an area that he or she is passionate about, and explore that interest through research, reading and writing in the coming weeks.

The unit culminated in a sixth-grade passion project conference, which took place after the conclusion of the study.

The newly developed passion project unit deviated from previous implementations of iPersonalize in two ways that are relevant to interpreting the results of the study:

1. The passion projects were not directly linked to the ELA standards. In effect, two units were running in parallel: one focused on helping students identify and explore

---

responses could not be linked to individual students or classrooms. In January 2018, FSD administered the survey a second time to students in the study.

[2] When FSD administered the survey, two teachers in the control group declined to administer the survey if the data were linked to individual students. For these two classrooms, individual student responses were obtained without student identification numbers. For the remaining classrooms, survey data were collected with student identification numbers.

their interests and strengths, and one focused on ELA standards such as word choice, supporting evidence, and pronouns.
2. The passion projects were not carried out in the context of an epic storyline or quest.

Furthermore, due to demands from the teachers union, the treatment and control groups both implemented most of the features of the iPersonalize approach, including using online assessments and other activities (via Thrively) and allowing student choice in activities. Table 2 provides information about how the game elements were incorporated into the iPersonalize unit implemented during the study in both the treatment and control groups. Implementation in the two groups was similar, with only the following differences:

- Only the treatment group had access to the online learning management system. The control group completed assessments and activities on other online platforms.
- Students in the treatment group earned online badges, while students in the control group received tangible rewards at classroom events.

Many of the activities in the instructional cycles were traditional classroom activities such as reading a passage and completing a worksheet, using a graphic organizer to complete a classifying activity, or viewing a video. Teacher materials reviewed for the study showed that teachers engaged students in a variety of learning formats—including large groups, small groups, pairs, and individuals—that followed FSD's adopted instructional cycle.

Table 2. Alignment of iPersonalize to Computer Game Elements (Landers, 2015)

Game Element	Elements of iPersonalize	Treatment	Control
Action language	Students are called *Agents of Change*.	X	X
	Students join teams and carry out team quests.	X	X
	Students participate in *Show What You Know* quests—online common formative assessments (CFAs).	X	X
Assessment	Students participate in *Show What You Know* Quests—online CFAs.	X	X
	Students take CFAs through the iPersonalize online learning management system.	X	
	Students take CFAs through another online platform.		X
	Students earn virtual badges online for mastering goals.	X	
	Students receive rewards (e.g., spirit sticks) at classroom celebrations.		X

Game Element	Elements of iPersonalize	Treatment	Control
	There is no summative assessment.	X	X
Challenges	There are secret passage games.	X	X
	There are bonus quests.	X	X
	In the *Deepening and Practicing* phase, students communicate what they learned and teach others, using a method of their choice such as a video blog (vlog), blog, or movie.	X	X
	There is no summative assessment for the cycles.	X	X
Control	In the *Deepening and Practicing* phase, students have options for products they can create: a vlog, blog, visual representation, movie, slide show, podcast, poem, social media post, song, or another option of their choosing.	X	X
	In the *Celebrate Success* phase, students receive badges for mastering the learning goals.	X	
Environment	Students used Thrively, an online system that supports students in discovering their strengths, interests, and aspirations through assessments and journals.	X	X
	Students log on to the iPersonalize online learning management system.	X	
	Students log on to another online platform.		X
Epic storyline	The passion project unit implemented for the study does not have a real epic storyline.	X	X
Interaction	Students work in large groups, small groups, and pairs, and as individuals, in activities throughout the lesson cycles.	X	X
	Students engage in feedback and peer review and provide GLOW comments (positive feedback) and GROW comments (to help other students do better next time).	X	X
Immersion	Immersion is not implemented in this unit.	X	X
Rules/goals	Instruction follows a defined instructional cycle (see Figure 1).	X	X

Game Element	Elements of iPersonalize	Treatment	Control
	Each cycle has consistent terminology (e.g., learning goals, a driving question, Bloom's taxonomy, GLOW, GROW, *Show What You Know*).	X	X

RESULTS

This section presents the results of the analyses that address each research question for the primary outcome variables: overall reading scale score; holistic writing score; and student engagement. Details on model specification are included in Appendix A. Descriptive statistics for all study measures are in Appendix B. Detailed results from all models are in Appendix C.

Prior to analyses to address the research questions, an examination of whether the treatment and control groups were equivalent at baseline was conducted. At pretest, students in the two groups scored similarly on both the reading scale score and the writing holistic score (see the *Baseline Equivalence* section in Appendix A for more details).

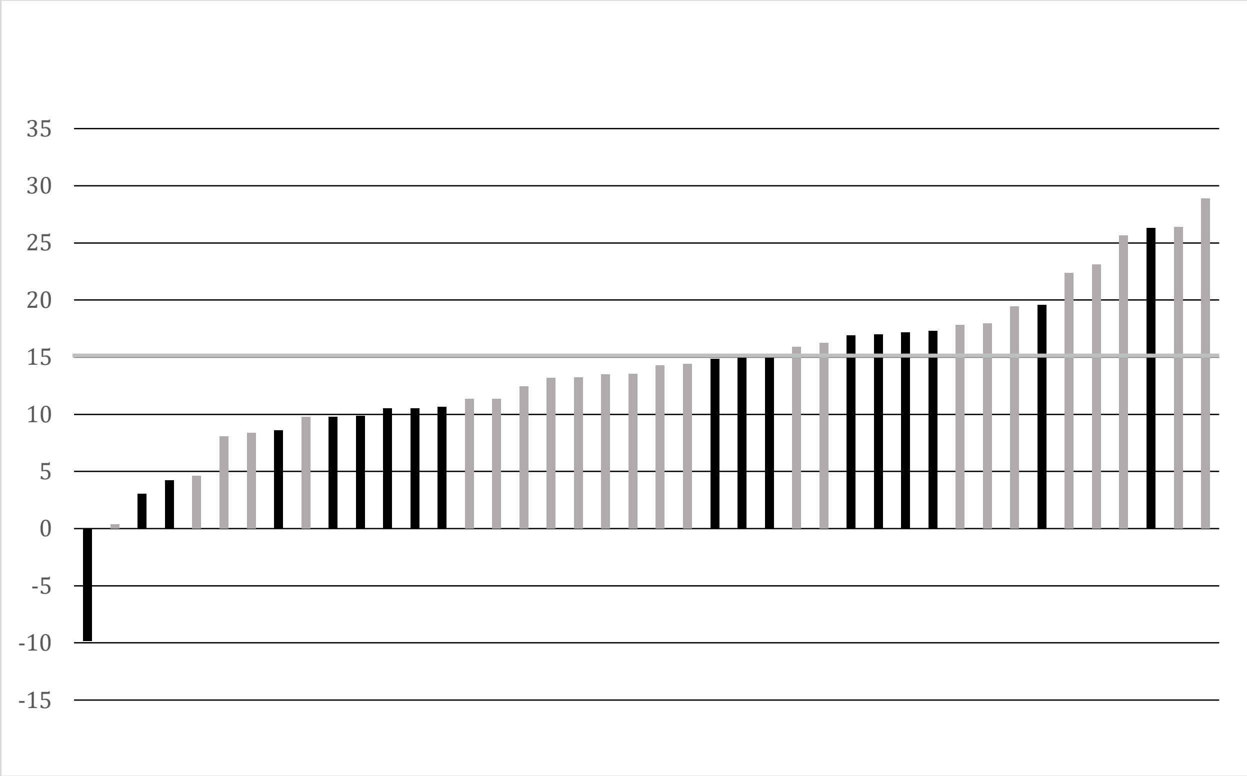
Research Question 1—What Is the Impact of iPersonalize on Average Sixth-Grade Student Achievement in ELA?

To address this question, multilevel models were fit. These models estimated the impact of iPersonalize on student reading scale scores and holistic writing scores after controlling for a range of student demographic characteristics, student pretest scores, and teacher characteristics. For both the reading scale score and the writing holistic score, the difference between the treatment group and control group was small and not statistically significant (*ns*), indicating that the two groups performed similarly:

- **Reading**: Effect of treatment = -.26, *ns*, effect size (ES) = -.01
- **Writing**: Effect of treatment = .00, *ns*, ES = .00 (see Appendix C, Tables C1 and C2)

The publishers of iReady have developed guidelines about the amount of change that should be expected during one year: expected growth in sixth grade is between 15 and 23 points (Curriculum Associates, 2017). With this expectation in mind, it is useful to consider the practical significance of the findings as well as the statistical significance. Figure 2 displays the average change in student reading scale score from pre- to post-test. Black bars represent classrooms assigned to the control condition, and gray bars represent classrooms assigned to the iPersonalize or treatment condition. In only the first trimester of the year, average student test scores improved by at least 15 points in 18 of the 42 classrooms. Ten of these classrooms were treatment classrooms, and eight were control classrooms. Overall, the proportion of treatment and control classrooms whose students improved their scores by more than 15 points was similar: 44 percent of control classrooms and 42 percent of treatment classrooms. A great deal of variability in average improvement in reading scale scores was evident for both the treatment and control classrooms.

Figure 2. Average Change in Reading Scale Scores from Pretest to Post-Test, by Classroom Type



Note. Black bars represent control classrooms; gray bars represent treatment classrooms.

Research Question 2—What Is the Impact of iPersonalize on Average Sixth-Grade Student Engagement in School?

To address this question, a model similar to the models fit to address research question 1 was used. The main difference was that, because the student engagement survey was administered only once, no pretest was included in the model. In addition, student responses to the survey in two classrooms could not be linked to other sources of data. As a result, these classrooms were not included in the model. Results indicated that students in treatment classrooms and students in control classrooms reported similar levels of engagement in school: effect of treatment = .05, *ns*, ES = .21 (see Appendix C, Table C3).

To guard against any bias that might have been introduced by eliminating two classrooms in the model above, an additional model was also run with the classroom average for engagement as the dependent variable. Using the classroom average as the outcome in this model allowed all 42 classrooms to be used. The same classroom averages and teacher experience variables were used as covariates in this model. Results were similar to the above model using student-level data as the outcome: effect of treatment = .04, *ns* (see

Appendix C, Table C4). Both treatment and control classrooms were similar in terms of students' reported engagement in school.

Research Question 3—To What Extent Does Gender Moderate the Impact of iPersonalize on Student Achievement or Student Engagement?

To address this question, an interaction term was added to test whether the effect of the intervention differed significantly for boys and girls.[3] For both the reading scale score and the writing holistic score, this effect was small and not statistically significant, indicating that the effect of the treatment was similar for boys and girls:

- **Reading scale score**: Treatment by gender effect = .23, *ns*, ES = .00
- **Writing holistic score**: Treatment by gender effect = -.12, *ns*, ES = -.06 (see Appendix C, Tables C5 and C6)

For student engagement, the effect size for the interaction was large enough to be considered substantively meaningful, although still statistically nonsignificant: treatment by gender effect = -.03, *ns*, ES = -.26 (see Appendix C, Table C7). The direction of the effect was negative, indicating that participating in iPersonalize was more strongly associated with student engagement for boys than it was for girls. To shed more light on this effect, a second set of models was fit, examining the effect of iPersonalize separately for boys and girls. These analyses revealed that boys in the treatment group reported higher levels of engagement in school than did boys in the control group, although the difference was not statistically significant: effect of treatment = .06, *ns*, ES = .22 (see Appendix C, Table C8). However, there was virtually no difference in student engagement between girls in the treatment group and girls in the control group: effect of treatment = .00, *ns*, ES = .01 (see Appendix C, Table C9).

Research Question 4—Is the Extent of Use of iPersonalize Associated with Increased Student Achievement?

To address this question, the study examined four variables assessing the extent to which students interacted with the iPersonalize online learning management system: the number of times each student logged on to the system; the number of quests each student attempted; the number of quests each student completed; and the percentage of attempted quests each student completed. To improve interpretability, the percentage of attempted quests was rescaled so that it ranged from 0 to 10. Thus, a one-unit change in the variable represents a 10 percent change in the attempted quests that were completed (e.g., a

[3] In all models, gender was coded 1 for girls and 0 for boys. Thus, a positive interaction term would indicate that the effect of iPersonalize was greater for girls than it was for boys. A negative interaction term would indicate that the effect of iPersonalize was greater for boys than it was for girls.

student who completed 80 percent of attempted quests would have a score of 8 on this variable).

Two sets of analyses were conducted to address research question 4. First, using only data from students assigned to the treatment group, the association between each of the usage variables and the two primary academic outcome variables (reading scale score and holistic writing score) was estimated (in separate models). All models controlled for a range of student demographic characteristics and student pretest scores.

The results of these models are presented in Table 3. None of the usage variables was associated with student performance on the writing holistic score. In reading, one usage variable—the percentage of attempted quests that were completed, which might be considered a measure of student persistence in completing quests—was significantly and positively associated with student performance. A 10 percent increase in the number of attempted quests that a student completed was associated with a 2.5 point increase on the reading scale score, which corresponds to about 17 percent of the expected growth for the sixth-grade year.

Table 3. Summary of Results of Models Examining the Association Between Student Usage Variables and Academic Outcomes

Usage Variable	Effect	Significance	Effect Size
Dependent variable: reading scale score			
Number of log-ons	-.01	*ns*	.00
Number of quests attempted	-.11	*ns*	.00
Number of quests completed	.23	*ns*	.00
Percentage of attempted quests that were completed	2.45	$p < .001$	.05
Dependent variable: holistic writing score			
Number of log-ons	.00	*ns*	.00
Number of quests attempted	.00	*ns*	.00
Number of quests completed	.00	*ns*	.00
Percentage of attempted quests that were completed	.04	*ns*	.04

Note. For complete results, see Appendix C, Tables C10–C17.

While the associations in Table 3 describe the relationship for the treatment group, they do not estimate the impact of increased usage on student achievement. This issue arose because the amount of usage was not randomly assigned to students, and it is likely associated with other unmeasured student characteristics that are correlated with both usage and achievement outcomes. To address this issue, a second set of models was fit, using data from students from both the treatment and control groups. Two techniques that rely on different assumptions to estimate the impact of increased usage on student achievement were applied. Both techniques are described in detail in Appendix A. For the reading and writing outcomes, the results of both sets of models indicated that the average causal effect of increases in usage was close to zero. Complete results from these models are presented in Appendix C, Table C18.

Research Question 5—What Is the Impact of iPersonalize on Low and High Quantiles of Sixth-Grade Student Achievement in ELA and Student Engagement in School?

All the regression models described thus far estimated average effects across the entire sample. Some descriptive statistics suggested that the impact of iPersonalize might be different for students at different percentiles in the achievement distribution. To formally explore this possibility, linear mixed-effects quantile regression models were fit (Geraci, 2014). These models estimate treatment effects at 10 different points along the achievement distribution (i.e., deciles). A detailed description of these models is provided in Appendix A.

Results indicated that the treatment effect on reading outcomes was very small (virtually zero) and consistent across the quantiles. On the other hand, for the writing outcome, there was more variance in the average treatment effect across the quantiles. At most quantiles, the average treatment effect on writing was very close to zero. However, larger estimates were found at higher quantiles, with the highest estimates at the 80th and 90th percentiles, corresponding to effect sizes of .19 and .20, respectively, which would suggest larger impacts for higher achieving students in writing (see Appendix C, Table C19). However, confidence intervals were fairly wide, containing zero in both cases. These larger impacts may simply be the result of random noise.

Consistent with the results of the mean regressions, there was somewhat more evidence of positive average treatment effects for student engagement relative to academic outcomes. Similar to the results for the writing outcome, the estimated average treatment effect was slightly larger at higher quantiles, although confidence intervals contained zero for all quantiles except for .90. For this quantile, the estimated effect was 10 times larger than the estimate at any other quantile (see Appendix C, Table C20). This result suggests a strong impact on engagement among the students who were already the most engaged.

CONCLUSIONS

This study was designed to examine the impact of iPersonalize on student achievement in ELA and student engagement in school. Unfortunately, issues with implementation of iPersonalize hampered the ability of the study to detect effects of the intervention. The main issue was substantial contamination of the control condition, which resulted in most students assigned to the control group receiving most of the intervention. Because the differences in the educational experience of treatment and control students was negligible, it is not surprising that there were no discernable differences in academic achievement between students assigned to the iPersonalize group and those assigned to business-as-usual instruction. However, students in both groups made gains on the reading and writing assessments.

In addition, the ability to detect improved student engagement was limited in this study. Timing the survey administration after the winter holiday break may have underestimated the true effect of the program on student engagement. Even so, the effect on engagement was the largest of the outcomes tested, and results suggested that the effect of iPersonalize for boys might be somewhat larger than its effect for girls. In addition, some evidence suggested that the impact on engagement might be most pronounced among students who were already the most engaged.

When the relationship between engagement with the iPersonalize online learning management system and student outcomes was examined, the results were mixed. Descriptive analyses of data from only the treatment group suggested that students who were more persistent in completing quests made greater gains in reading. However, more rigorous analyses that used data from students in both groups indicated that this relationship is most likely not causal.

LESSONS LEARNED

This study was funded by a low-cost, short-duration grant from the Institute of Education Sciences (IES). These grants are designed to provide a way for state and local education agencies to obtain rigorous evidence of the effectiveness of educational interventions in a relatively short period of time (IES, 2018). These grants have very clear guidelines about how the funds may be used. Specifically, funds cannot be used to support implementation of a program or collection of data, which limits the ability of researchers to be involved in these critical components of a study.

These grants also require a partnership between researchers and an education agency. For this project, Marzano Research conceptualized our partnership as a *district-focused research alliance* (Coburn, Penuel, & Geil, 2013). In such a partnership, intense collaboration at the beginning of a project aims to identify the problem to be studied and determine the best approach to addressing the research questions. While the study and

analyses are being conducted, the two parties are relatively independent so that the objectivity of the research is not compromised. Marzano Research planned for a district-focused research alliance because it seemed particularly compatible with the requirements of the grant (i.e., limiting researcher involvement in implementation and data collection).

In the end, this approach to the partnership was not successful. During the period in which FSD and the researchers were operating fairly independently, decisions related to implementation were made without the involvement of the research team. These decisions had an important impact on the conclusions that could be drawn from the study. In future research with this type of funding, researchers should maintain closer involvement so that they can monitor implementation more closely, within the guidelines of the grant. Closer involvement could include very frequent communication about how implementation is progressing and, possibly, site visits to observe implementation.

This study would have also benefitted from a longer planning period with greater collaboration between FSD and the researchers prior to the implementation of the intervention. A low-cost, short-duration grant cannot accommodate a lengthy planning period, but this study might have been more successful if we had planned for implementation in the spring of the school year and used the fall as a planning period.

Finally, iPersonalize was possibly not a good candidate for a study conducted with this type of funding. It is a complex and multifaceted intervention. In addition, FSD was still in the process of refining iPersonalize, as evidenced by the implementation of a new unit during the study. In hindsight, iPersonalize would have been a better candidate for study with a grant that allowed for iterative development and pilot-testing of the intervention, such as IES (2019) development and innovation grants. Low-cost, short-duration grants might be a better funding mechanism to support less complex and more established interventions.

APPENDIX A. STUDY DESIGN AND RESEARCH METHODS

This appendix provides additional technical details about the study design and research methods used in the iPersonalize evaluation.

Random Assignment of Teachers

Because iPersonalize involves instruction delivered at the classroom level, the unit of randomization was the teacher. Several factors were considered when determining the optimal time to randomize teachers to condition. Randomizing early enough that teachers could prepare to effectively implement iPersonalize was of primary importance. FSD holds an annual training on iPersonalize in June. Therefore, random assignment had to occur prior to the training. Initially, the plan was to randomly assign teachers to condition after preliminary student class lists were determined in May. However, this plan turned out not to be feasible, as class lists would not be determined until after the training in June. As a result, students were assigned to teachers after teachers were randomly assigned to condition.

In May 2017, after FSD made final staffing decisions for 2017/18, the district informed all sixth-grade teachers of the purpose of the study and the chance that they could be assigned to either the treatment or control condition. Forty teachers from 15 schools volunteered to participate. The number of teachers per school ranged from one to five. FSD forwarded a list of teachers who had volunteered to participate in the study to Marzano Research. Using a random number generator, teachers were randomly assigned, stratifying by school, to the treatment condition (i.e., classrooms delivering iPersonalize in ELA during first trimester of the 2017/18 school year) or to the control condition (i.e., classrooms providing business-as-usual instruction during the first trimester of the year). If only one teacher in a school volunteered, that teacher was randomly assigned to one condition, using a random number generator set to yield values of only 1 (treatment) or 0 (control). Nineteen teachers were assigned to the control group, and 21 teachers were assigned to the treatment group (Table A1).

After assignment, but before the start of the school year, four teachers assigned to the control condition left their positions as sixth-grade teachers. One of these teachers was replaced with a new teacher who was willing to participate in the study. This teacher was assigned to the same condition as the teacher replaced.

Two of the participating schools implement an instructional model in which one teacher provides ELA instruction to all sixth-graders at the school. In one of these schools, one participating teacher was assigned to the control condition and taught three groups of students. In the other school, one participating teacher was assigned to the treatment condition and taught four groups of students. Each of these groups of students was treated

as a distinct classroom in the analyses. The addition of five classrooms resulted in a total of 42 classrooms for the study—24 treatment and 18 control (Table A1).

In another school, two teachers who were initially assigned to different conditions worked together as a team. The teacher initially assigned to the treatment condition provided ELA instruction to both classes of students. This teacher maintained the separate group assignment for each class. That is, the teacher provided instruction with iPersonalize to her own class while providing business-as-usual instruction to her teaching partner's class.[4]

Finally, one teacher who was assigned to the treatment group did not actually implement the program. In keeping with the intent-to-treat approach, this teacher and her students were considered members of the treatment group in all analyses.

Table A1. Number of Classrooms, by Condition

Classroom Description	Treatment	Control
Teachers initially randomized	21	19
Teachers who left the study and were not replaced	0	3
Additional groups of students taught by teachers	3	2
Total classrooms in final sample	24	18

After the pretest assessment was administered in September 2017, the Technology and Media Services Department at FSD sent email invitations to students in the treatment group, granting them access to iPersonalize. Invitations were not sent to students in the control group, ensuring that these students did not have access to the iPersonalize online learning management system during the study.

Data Analysis Procedures

To address the two primary research questions about the impact of the iPersonalize program on student achievement in ELA (research question 1) and student engagement (research question 2), the study used hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002). HLM was well suited to this study due to a mismatch between the unit of assignment to condition (i.e., teacher) and the unit at which the outcome of interest was measured (i.e., student). Students with missing data on the pretest, post-test, or covariates

[4] Students in her control classroom did not have access to the iPersonalize online learning management system.

were eliminated from the analyses.[5] A series of two-level models were fit. The models included students at level 1 and teachers/classrooms at level 2.

Baseline Equivalence

The first set of models examined baseline equivalence of the two groups, using the pretest benchmark test scores (i-Ready Diagnostic or MY Access! School Edition). Only students who had data for both pretest and post-test scores and all covariates were included to ensure that the analytic samples used in the baseline equivalence and full models were comparable. The statistical significance of the estimate of the baseline difference between the treatment and control group was examined, and an effect size was computed (Cohen's *d*). Analyses yielded effect sizes of .07 and .05 for pretest reading and writing scores, respectively, which are rather small. Even though the groups did not differ substantially at baseline, all impact analyses included baseline assessment scores as covariates.

Analyses to Address Research Questions

To address research question 1, a second set of models was fit. This set of models provided an estimate of the impact of iPersonalize on ELA achievement for sixth-grade students. At level 1, these models included the pretest benchmark score (i-Ready or MY Access!), student gender, free or reduced-price lunch status, English learner status, gifted status, and dummy variables for race/ethnicity as fixed effects. All of these level-1 predictors were group-mean centered by subtracting the classroom mean for the variable from each student's score. At level 2, a dummy-coded variable representing treatment status was included as an independent variable along with a set of classroom-level covariates, including teacher characteristics (e.g., years of experience teaching, master's degree or no

---

[5] For the analysis predicting reading scale score, 48 students were dropped due to missing data. Twenty-five of these students were from the control group and their patterns of missing data were as follows: missing only the pretest, 4; missing only the posttest, 11; missing a covariate, 6; missing pretest and a covariate, 1; missing posttest and a covariate, 2; and missing pretest, posttest, and a covariate, 1. Twenty-three of these students were from the iPersonalize group and their patterns of missing data were as follows: missing only the pretest, 5; missing only the posttest, 12; missing a covariate, 3; missing pretest and posttest, 2; and missing posttest and a covariate, 1. For the analysis predicting the writing holistic score, 144 students were dropped due to missing data. Eighty-one of these students were from the control group and their patterns of missing data were as follows: missing only the pretest, 22; missing only the posttest, 33; missing a covariate, 5; missing posttest and a covariate, 2; missing pretest and posttest, 16; and missing pretest, posttest, and a covariate, 3. Sixty-three of these students were from the iPersonalize group and their patterns of missing data were as follows: missing only the pretest, 29; missing only the posttest, 22; missing a covariate, 1; missing pretest and a covariate, 2; missing posttest and a covariate, 1; and missing pretest and posttest, 8. For the analysis predicting student engagement, 158 students were dropped due to missing data. One hundred two of these students were in the control group. Sixty-one were enrolled in the two classrooms where the teachers declined to administer the survey with student identification numbers. From the remaining control classrooms, 33 students were dropped because they were missing the YouthTruth survey, 4 were dropped because they were missing a covariate, and 4 were dropped because they were missing the YouthTruth survey and a covariate. Fifty-six students were dropped from the iPersonalize group and their patterns of missing data were as follows: missing YouthTruth survey, 52; missing a covariate, 1; and missing both the YouthTruth survey and a covariate, 3.

master's degree, and years of experience implementing iPersonalize), dummy codes for school, classroom average on the pretest benchmark score, and classroom averages for the same student demographic characteristics included at level 1. All covariates at level 2 were grand-mean centered by subtracting the overall sample mean for the variable from each classroom's score. The statistical significance of the effect of treatment status along with the effect size were examined. Following What Works Clearinghouse (2017) guidance, an effect size of at least 0.25 was considered substantively meaningful, even if it was not statistically significant.

The model fit to address research question 2 was similar to the models fit for research question 1. The main difference was that student engagement was the outcome variable. Because there was no pretest for this variable, no pretest covariate was included in the model.

To address research question 3, a second equation was estimated at level 2. This equation included treatment as a predictor of the level-1 slope for gender. By specifying this additional equation, the model included a cross-level interaction term that estimated the moderating effect of gender on the average treatment effect. To aid in interpretability, gender was not centered at level 1, and the classroom-level mean for gender was not included in the model. In addition, follow-up models were fit separately for each gender. The overall treatment effect of iPersonalize for each gender separately was of primary interest in these models.

Two sets of analyses were conducted to address research question 4. The first set of models included only students from the treatment group. Because, by definition, treatment status would no longer vary, this variable was dropped from the level-2 model. Using four separate models, each of the four iPersonalize usage variables was added as a predictor at level 1. The coefficients for the usage predictor in each model was examined to determine the extent to which the relevant feature of iPersonalize usage was associated with student outcomes (ELA achievement and engagement). Because only the 24 treatment classrooms were included in these analyses, the models had fewer degrees of freedom. Thus, the list of covariates in the models was reduced. The two smallest racial/ethnic minority groups were combined (White and Other) and a single dummy variable was used for the combined group, and the teacher experience variables were not included as covariates.

The second set of analyses for research question 4 included students from both groups and provide better estimates of the causal relationship between usage and impacts. Two techniques that rely on different assumptions to estimate the impact of increased usage on student achievement were applied. The first technique, instrumental variables (see, for example, Gennetian, Morris, Bos, & Bloom, 2005), treats the randomization indicator as an instrument for the relevant usage variable. This technique relies on the assumption that the only way that randomization can impact outcomes is through its impact on the relevant usage variable. While this assumption obviously could not be true across all four usage variables simultaneously, numerical results for all four usage variables are still presented

here to let the reader evaluate the possible validity of the assumptions. As described in Angrist and Krueger (2001), this technique identifies a *local average treatment effect* (LATE) for those whose behavior was impacted by being assigned to an iPersonalize classroom. This LATE may not represent the average treatment effect for the entire population of students.

The instrumental variable models were estimated using two-stage least squares (2SLS) methods as implemented by the ivreg function in the R package AER (Kleiber & Zeileis, 2008). Cluster robust standard errors were computed using the cluster.robust.se function in the R package ivpack (Jiang & Small, 2015). Eight models were fit using each technique and model specification, one for each combination of outcome variable and usage variable. These analyses used the 1,054 students with complete data for pretest and posttest for both outcome variables.

The initial specification of the 2SLS models included the usage variable as an endogenous variable, the randomization indicator as an instrument, and pretests as additional exogenous variables. A second specification added all other covariates as exogenous variables, namely the level-1 demographic predictors listed above, indicator variables for school membership, and level-2 teacher experience and qualification variables. All variables were uncentered.

The second technique, a continuous version of the Analysis of Symmetrically Predicted Endogenous Subgroups (ASPES) method first proposed in Peck (2003), is new to the literature, having been discussed only briefly in Peck (2015). This technique relies on baseline, pretreatment covariates, which are used to predict the usage variable with only treatment group data. This regression model is employed to create predicted usage values in both the treatment and control groups. The predicted usage values are then used as an independent variable in a regression model to predict the outcome of interest. The regression model also contains an indicator variable for randomization and the interaction between the randomization indicator and the predicted usage variable. Because the predicted values depend only on pretreatment variables, the estimated coefficients do not suffer from the endogeneity problem that exists for the values reported in Table 3 above. The coefficient on the interaction term shows how variation in usage relates to variation in intervention impact. The coefficients for these interaction terms are reported in Table C18 . As noted in Peck (2015), these coefficients can be interpreted as applying to actual rather than predicted usage values, provided that impacts conditional on the predicted usage variable are the same as impacts conditional on the actual usage variable.

The process for fitting continuous ASPES models was as follows. To avoid overfitting, a cross-validation approach was used (Moulton, Peck, & Greeney, 2018). Stratified random sampling was used to create 10 random subsets of the data. Each subset contained 56 or 57 treatment cases and 48 or 49 control cases. For each subset i = 1, …, 10, the treatment group data for all subsets besides i were used to create a regression model predicting the usage variable from either (model 1) all pretreatment covariates (namely those covariates

listed in the above paragraphs) or (model 2) all pretreatment covariates *except* the pretests. All variables were uncentered. The regression models were random intercept HLM models with students nested within classrooms, and they were fit using the lmer function in the lme4 package in R (Bates, Machler, Bolker, & Walker, 2015). In some cases, certain school indicator variables needed to be removed from the models because a particular school was not represented in a subset of the data. Using the parameter estimates from these models, predicted usage values were created for all treatment *and* control cases in subset *i*. This process was repeated for each of the four usage variables.

Once the predicted values were obtained, random intercept HLM models were fit with one of the two outcome variables (reading or writing) as the dependent variable. The predicted usage variable, the randomization indicator, the interaction between predicted usage and the randomization indicator, and either (model 1) all pretreatment covariates or (model 2) pretests (for a total of eight models across the four usage variables) were used as independent variables. The coefficients and associated *p* values for these interaction terms (reported in Appendix C, Table C18) are those associated with the model 2 specification. The *p* values were found using a Satterthwaite approximation via the R package lmerTest (Kuznetsova, Brockhoff, & Christensen, 2019). While the significance levels reported do not account for estimation error in the first stage regressions, previous explorations with the discrete ASPES approach indicated that the contribution of the first-stage estimation error to the overall error was trivial (L. Peck, personal communication, July 22, 2019).

To address research question 5, the R package lqmm (Geraci, 2014) was used to fit linear quantile mixed models. For each of the achievement outcomes of interest, treatment effects at the deciles of the achievement distribution were estimated. Ninety-five percent confidence intervals around these estimates are computed. Four sets of models of increasingly complexity were fit. The first model type had only an indicator for treatment assignment as an independent variable. The second model type added fixed effects for school membership. The third model type added the subject-specific pretest. The fourth model type also included student-level demographic covariates (uncentered). In all cases, random intercept models were fit. Teacher-level variables were not included in the models because their inclusion often caused the fitting algorithm to fail to converge and otherwise increased standard errors. The results from the fourth model type are presented in Appendix C, Table C19. Results for the reading outcome are based on the 1,247 students with complete pretest, post-test, and demographic information. Results for the writing outcome are based on the 1,151 students with complete pretest, post-test, and demographic information. While numerical results were somewhat sensitive to the covariate set included in the models, substantive interpretations would have remained consistent had the other models that were fit been presented.

The lqmm package fits quantile regression models while accounting for a nondiagonal covariance matrix by maximizing a likelihood function defined in terms of the asymmetric Laplace distribution. The likelihood is maximized using a modified Gaussian quadrature

approach. The default settings in the lqmm package were used to fit the models, corresponding to a Gauss-Hermite quadrature approach with seven nodes.

Confidence intervals were obtained using the block bootstrap approach described in Geraci and Bottai (2014) and implemented via the summary() function in lqmm. The default of 50 bootstrap replications was used.

APPENDIX B. DESCRIPTIVE STATISTICS FOR STUDY MEASURES

Table B1. Descriptive Statistics for Academic and Engagement Outcomes

Variable	*N*	Mean	*SD*	Minimum	Maximum
Reading scale score—pretest	1,256	574.78	55.03	317.00	727.00
Reading scale score—post-test	1,256	588.45	52.01	359.00	727.00
Writing holistic score—pretest	1,157	3.00	0.98	1.00	6.00
Writing holistic score—post-test	1,157	3.73	1.08	1.00	6.00
YouthTruth score	1,210	2.48	0.26	1.24	3.00

Table B2. Descriptive Statistics for iPersonalize Usage Variables (treatment group only)

Variable	*N*	Mean	*SD*	Minimum	Maximum
Number of log-ons	607	106.35	116.73	5	891
Number of attempted quests	606	21.96	7.16	4	40
Number of completed quests	606	17.76	7.66	2	40
Percent of attempted quests that were completed	606	79.10	17.24	30	100

APPENDIX C. DETAILED RESULTS FOR STATISTICAL MODELS

This appendix provides tables summarizing the coefficients for all variables in the models that were used to address the research questions for the study. Results are presented by research question.

Research Question 1—What Is the Impact of iPersonalize on Average Sixth-Grade Student Achievement in ELA?

Two HLM models were fit to address research question 1. One model included the reading scale score as the dependent variable (Table C1), and one model included the writing holistic score as the dependent variable (Table C2).

Table C1. HLM Coefficients for Model Predicting Reading Scale Score (*n* = 42 classrooms, 1,247 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	587.22	1.30	450.00	14	< 0.001
Treatment	-0.26	1.93	-0.13	14	0.90
Percentage of female students	12.45	25.19	0.49	14	0.63
Percentage of students qualifying for free or reduced-price lunch	6.72	21.29	0.32	14	0.76
Percentage of students identified as English learners	-8.04	13.29	-0.61	14	0.56
Percentage of students qualifying for gifted services	6.89	7.68	0.90	14	0.39
Percentage of students receiving special education services	15.88	21.74	0.73	14	0.48
Percentage Hispanic	8.47	15.22	0.56	14	0.59
Percentage Asian/Pacific Islander	15.14	16.80	0.90	14	0.38
Percentage Other race/ethnicity mean	45.90	34.94	1.31	14	0.21
Pretest reading scale score classroom mean	0.81	0.09	9.12	14	< 0.001
School 1	10.11	10.22	0.99	14	0.34
School 2	4.52	14.88	0.30	14	0.77
School 3	6.13	6.17	0.99	14	0.34
School 4	3.36	15.98	0.21	14	0.84

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
School 5	6.53	6.99	0.93	14	0.37
School 6	1.50	9.09	0.17	14	0.87
School 7	8.92	14.38	0.62	14	0.55
School 8	1.35	6.94	0.19	14	0.85
School 9	3.24	5.71	0.57	14	0.58
School 10	-4.08	5.29	-0.77	14	0.45
School 11	2.16	5.55	0.39	14	0.70
School 12	-10.82	12.82	-0.84	14	0.41
School 13	1.07	14.95	0.07	14	0.94
School 14	7.29	5.27	1.39	14	0.19
Teacher with master's degree	1.80	2.80	0.64	14	0.53
Teacher with prior experience with iPersonalize	5.36	3.56	1.51	14	0.15
Years of teaching experience	0.07	0.13	0.51	14	0.62
Student Level					
Pretest reading scale score	0.75	0.02	45.31	1,196	< 0.001
Female	2.76	1.28	2.16	1,196	0.03
Qualifying for free or reduced-price lunch	-6.08	1.84	-3.30	1,196	< 0.01
English learner	-6.39	1.88	-3.39	1,196	< 0.001
Qualifying for gifted services	0.91	2.96	0.31	1,196	0.76
Receiving special education services	-1.19	2.14	-0.55	1,196	0.58
Hispanic	-3.26	2.32	-1.41	1,196	0.16
Asian/Pacific Islander	2.48	2.32	1.07	1,196	0.29
Other race/ethnicity	-0.58	3.89	-0.15	1,196	0.88

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C2. HLM Coefficients for Model Predicting Writing Holistic Score (*n* = 42 classrooms, 1,151 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	3.68	0.10	36.05	14	< 0.001
Treatment	0.00	0.16	0.02	14	0.99
Percentage of female students	-1.60	2.09	-0.77	14	0.46
Percentage of students qualifying for free or reduced-price lunch	2.40	1.80	1.33	14	0.20
Percentage of students identified as English learners	-2.20	0.79	-2.79	14	0.02
Percentage of students qualifying for gifted services	-0.52	0.49	-1.07	14	0.30
Percentage of students receiving special education services	-0.29	1.63	-0.18	14	0.86
Percentage Hispanic	-0.24	1.16	-0.21	14	0.84
Percentage Asian/Pacific Islander	2.23	1.26	1.76	14	0.10
Percentage Other race/ethnicity mean	2.42	2.64	0.92	14	0.37
Pretest writing holistic score classroom mean	0.61	0.23	2.62	14	0.02
School 1	0.50	0.91	0.55	14	0.59
School 2	-0.29	1.26	-0.23	14	0.82
School 3	-0.62	0.51	-1.21	14	0.25
School 4	0.28	1.17	0.24	14	0.81
School 5	0.37	0.66	0.56	14	0.59
School 6	0.21	0.97	0.22	14	0.83
School 7	0.37	1.17	0.31	14	0.76
School 8	0.14	0.59	0.24	14	0.81
School 9	0.23	0.43	0.54	14	0.60
School 10	-0.57	0.44	-1.31	14	0.21
School 11	0.15	0.46	0.34	14	0.74
School 12	0.11	1.13	0.09	14	0.93
School 13	-0.27	1.14	-0.24	14	0.81
School 14	0.56	0.46	1.21	14	0.25
Teacher with master’s degree	0.26	0.20	1.30	14	0.21

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Teacher with prior experience with iPersonalize	-0.08	0.26	-0.31	14	0.76
Years of teaching experience	0.01	0.01	0.68	14	0.51
Student Level					
Pretest writing holistic score	0.46	0.03	14.68	1,100	< 0.001
Female	0.00	0.05	-0.09	1,100	0.93
Qualifying for free or reduced-price lunch	-0.05	0.07	-0.74	1,100	0.46
English learner	-0.34	0.07	-4.95	1,100	< 0.001
Qualifying for gifted services	0.03	0.11	0.31	1,100	0.75
Receiving special education services	-0.26	0.08	-3.29	1,100	< 0.01
Hispanic	-0.05	0.08	-0.63	1,100	0.53
Asian/Pacific Islander	0.12	0.08	1.45	1,100	0.15
Other race/ethnicity	-0.11	0.14	-0.78	1,100	0.44

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Research Question 2—What Is the Impact of iPersonalize on Average Sixth-Grade Student Engagement in School?

One HLM model (Table C3) using student and classroom level data and one weighted least squares (WLS) regression model using only classroom level data (Table C4) were fit to address research question 2.

Table C3. HLM Coefficients for Model Predicting Student Engagement (*n* = 40 classrooms, 1,137 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	2.45	0.03	81.02	13	< 0.001
Treatment	0.05	0.04	1.22	13	0.25
Percentage of female students	0.45	0.47	0.96	13	0.36
Percentage of students qualifying for free or reduced-price lunch	0.31	0.39	0.79	13	0.44
Percentage of students identified as English learners	0.10	0.25	0.41	13	0.69

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Percentage of students qualifying for gifted services	0.01	0.11	0.13	13	0.90
Percentage of students receiving special education services	-0.46	0.45	-1.01	13	0.33
Percentage Hispanic	-0.31	0.27	-1.14	13	0.27
Percentage Asian/Pacific Islander	-0.10	0.29	-0.35	13	0.73
Percentage Other race/ethnicity mean	-0.93	0.64	-1.45	13	0.17
School 1	0.22	0.19	1.20	13	0.25
School 2	0.19	0.27	0.68	13	0.51
School 3	0.03	0.11	0.30	13	0.77
School 4	0.13	0.25	0.52	13	0.62
School 5	0.34	0.14	2.52	13	0.03
School 6	0.18	0.18	1.00	13	0.33
School 7	0.26	0.26	1.02	13	0.33
School 8	0.02	0.11	0.19	13	0.85
School 9	0.10	0.10	1.05	13	0.31
School 10	-0.15	0.10	-1.53	13	0.15
School 11	0.02	0.09	0.21	13	0.84
School 12	0.26	0.25	1.03	13	0.32
School 13	0.11	0.26	0.42	13	0.68
School 14	-0.24	0.33	-0.72	13	0.49
Teacher with master’s degree	-0.01	0.06	-0.24	13	0.82
Teacher with prior experience with iPersonalize	-0.04	0.07	-0.54	13	0.60
Years of teaching experience	0.00	0.00	0.77	13	0.46
Student Level					
Female	0.06	0.01	4.40	1,089	< 0.001
Qualifying for free or reduced-price lunch	-0.02	0.02	-1.21	1,089	0.23
English learner	0.01	0.02	0.33	1,089	0.75
Qualifying for gifted services	-0.01	0.03	-0.42	1,089	0.68
Receiving special education services	-0.03	0.02	-1.23	1,089	0.22
Hispanic	0.04	0.03	1.61	1,089	0.11
Asian/Pacific Islander	-0.04	0.03	-1.57	1,089	0.12

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Other race/ethnicity	0.07	0.04	1.64	1,089	0.10

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C4. WLS Regression Coefficients for Model Predicting Classroom-Level Engagement (*n* = 42 classrooms)

Predictor	Coefficient	Standard Error	*t* Ratio	*p* Value
Intercept	2.35	0.26	9.16	< 0.001
Treatment	0.04	0.04	1.20	0.25
Percentage of female students	0.50	0.43	1.16	0.26
Percentage of students qualifying for free or reduced-price lunch	0.21	0.36	0.58	0.57
Percentage of students identified as English learners	0.14	0.22	0.61	0.55
Percentage of students qualifying for gifted services	-0.05	0.12	-0.45	0.66
Percentage of students receiving special education services	-0.35	0.38	-0.93	0.34
Percentage Hispanic	-0.49	0.26	-1.84	0.09
Percentage Asian/Pacific Islander	0.02	0.30	0.06	0.96
Percentage Other race/ethnicity mean	-0.56	0.53	-1.07	0.30
School 1	0.09	0.17	0.49	0.63
School 2	-0.03	0.24	-0.12	0.91
School 3	0.03	0.10	0.31	0.76
School 4	-0.12	0.27	-0.44	0.67
School 5	0.24	0.12	2.05	0.06
School 6	0.09	0.15	0.62	0.54
School 7	0.01	0.25	0.03	0.98
School 8	0.08	0.11	0.76	0.46
School 9	0.10	0.09	1.11	0.29
School 10	-0.15	0.09	-1.68	0.11
School 11	0.03	0.09	0.30	0.77
School 12	0.10	0.23	0.42	0.68

Predictor	Coefficient	Standard Error	*t* Ratio	*p* Value
School 13	-0.10	0.26	-0.40	0.70
School 14	0.12	0.10	1.15	0.27
Teacher with master's degree	-0.02	0.05	-0.52	0.61
Teacher with prior experience with iPersonalize	-0.06	0.06	-0.89	0.39
Years of teaching experience	0.00	0.00	1.44	0.17

Note. Analysis was weighted by classroom size.

Research Question 3—To What Extent Does Gender Moderate the Impact of iPersonalize on Student Achievement or Student Engagement?

Research question 3 was first addressed with three HLM models. These models included reading scale score (Table C5), the writing holistic score (Table C6), and the YouthTruth score (Table C7) as dependent variables. For student engagement, the effect size for the interaction between treatment and gender was large enough to be considered substantively meaningful, although still statistically nonsignificant. To aid in interpreting this effect, separate follow-up models were fit for boys (Table C8) and girls (Table C9).

Table C5. HLM Coefficients for Model Predicting Reading Scale Score with Gender as a Moderating Variable (*n* = 42 classrooms, 1,247 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	585.89	1.60	367.11	15	< 0.001
Treatment	-0.39	2.28	-0.17	15	0.87
Percentage of students qualifying for free or reduced-price lunch	12.79	13.85	0.92	15	0.37
Percentage of students identified as English learners	-10.44	11.41	-0.92	15	0.38
Percentage of students qualifying for gifted services	6.41	7.39	0.87	15	0.40
Percentage of students receiving special education services	12.68	19.62	0.65	15	0.53
Percentage Hispanic	8.01	14.78	0.54	15	0.60
Percentage Asian/Pacific Islander	15.42	16.35	0.94	15	0.36
Percentage Other race/ethnicity mean	43.03	33.40	1.29	15	0.22

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Pretest reading scale score classroom mean	0.82	0.09	9.51	15	< 0.001
School 1	12.25	8.42	1.46	15	0.17
School 2	7.85	11.92	0.66	15	0.52
School 3	6.17	6.02	1.03	15	0.32
School 4	6.29	13.81	0.46	15	0.66
School 5	7.06	6.71	1.05	15	0.31
School 6	3.55	7.26	0.49	15	0.63
School 7	11.67	12.31	0.95	15	0.36
School 8	1.88	6.63	0.28	15	0.78
School 9	3.46	5.54	0.62	15	0.54
School 10	-4.06	5.16	-0.79	15	0.44
School 11	2.67	5.26	0.51	15	0.62
School 12	-7.76	9.90	-0.78	15	0.45
School 13	4.46	11.94	0.37	15	0.71
School 14	7.52	5.10	1.47	15	0.16
Teacher with master's degree	1.33	2.46	0.54	15	0.60
Teacher with prior experience with iPersonalize	5.82	3.29	1.77	15	0.10
Years of teaching experience	0.06	0.13	0.50	15	0.62
Student Level					
Pretest reading scale score	0.75	0.02	45.28	1,195	< 0.001
Female	2.66	1.91	1.39	1,195	0.16
Treatment by female interaction	0.23	2.54	0.09	1,195	0.93
Qualifying for free or reduced-price lunch	-6.08	1.84	-3.30	1,195	< 0.01
English learner	-6.39	1.89	-3.39	1,195	< 0.001
Qualifying for gifted services	0.92	2.97	0.31	1,195	0.76
Receiving special education services	-1.18	2.14	-0.55	1,195	0.58
Hispanic	-3.26	2.32	-1.40	1,195	0.16
Asian/Pacific Islander	2.48	2.32	1.07	1,195	0.29
Other race/ethnicity	-0.57	3.89	-0.15	1,195	0.88

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, female was uncentered, and all other predictors were centered on the classroom mean for the variable.

Table C6. HLM Coefficients for Model Predicting Writing Holistic Score with Gender as a Moderating Variable (*n* = 42 classrooms, 1,151 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	3.65	0.11	34.58	15	< 0.001
Treatment	0.05	0.16	0.31	15	0.76
Percentage of students qualifying for free or reduced-price lunch	1.32	1.08	1.22	15	0.24
Percentage of students identified as English learners	-1.92	0.68	-2.82	15	0.01
Percentage of students qualifying for gifted services	-0.43	0.46	-0.93	15	0.37
Percentage of students receiving special education services	-0.16	1.59	-0.10	15	0.92
Percentage Hispanic	0.16	1.03	0.15	15	0.88
Percentage Asian/Pacific Islander	2.21	1.24	1.78	15	0.10
Percentage Other race/ethnicity mean	3.12	2.46	1.27	15	0.22
Pretest writing holistic score classroom mean	0.57	0.22	2.56	15	0.02
School 1	0.05	0.66	0.07	15	0.94
School 2	-0.87	0.98	-0.89	15	0.39
School 3	-0.74	0.47	-1.57	15	0.14
School 4	-0.27	0.90	-0.30	15	0.77
School 5	0.14	0.57	0.24	15	0.81
School 6	-0.32	0.65	-0.50	15	0.63
School 7	-0.15	0.94	-0.16	15	0.88
School 8	-0.10	0.48	-0.22	15	0.83
School 9	0.13	0.40	0.32	15	0.75
School 10	-0.70	0.40	-1.74	15	0.10
School 11	-0.04	0.37	-0.12	15	0.91
School 12	-0.48	0.80	-0.60	15	0.56
School 13	-0.86	0.83	-1.05	15	0.31
School 14	0.36	0.38	0.97	15	0.35
Teacher with master's degree	0.30	0.19	1.56	15	0.14

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Teacher with prior experience with iPersonalize	-0.13	0.25	-0.51	15	0.62
Years of teaching experience	0.01	0.01	0.72	15	0.49
Student Level					
Pretest writing holistic score	0.46	0.03	14.70	1,099	< 0.001
Female	0.07	0.07	0.94	1,099	0.35
Treatment by female interaction	-0.12	0.09	-1.34	1,099	0.18
Qualifying for free or reduced-price lunch	-0.05	0.07	-0.71	1,099	0.48
English learner	-0.33	0.07	-4.94	1,099	< 0.001
Qualifying for gifted services	0.03	0.11	0.27	1,099	0.78
Receiving special education services	-0.27	0.08	-3.33	1,099	< 0.001
Hispanic	-0.05	0.08	-0.66	1,099	0.51
Asian/Pacific Islander	0.12	0.08	1.43	1,099	0.15
Other race/ethnicity	-0.12	0.14	-0.83	1,099	0.40

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, female was uncentered, and all other predictors were centered on the classroom mean for the variable.

Table C7. HLM Coefficients for Model Predicting YouthTruth Score with Gender as a Moderating Variable (*n* = 40 classrooms, 1,137 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	2.41	0.03	75.75	14	< 0.001
Treatment	0.07	0.05	1.47	14	0.17
Percentage of students qualifying for free or reduced-price lunch	0.55	0.27	2.03	14	0.06
Percentage of students identified as English learners	-0.01	0.21	-0.02	14	0.98
Percentage of students qualifying for gifted services	0.00	0.11	-0.03	14	0.98
Percentage of students receiving special education services	-0.58	0.42	-1.37	14	0.19
Percentage Hispanic	-0.33	0.27	-1.22	14	0.24

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Percentage Asian/Pacific Islander	-0.07	0.28	-0.26	14	0.80
Percentage Other race/ethnicity mean	-1.02	0.62	-1.64	14	0.12
School 1	0.31	0.16	1.98	14	0.07
School 2	0.32	0.22	1.45	14	0.17
School 3	0.04	0.11	0.33	14	0.75
School 4	0.24	0.21	1.11	14	0.29
School 5	0.36	0.13	2.69	14	0.02
School 6	0.26	0.15	1.75	14	0.10
School 7	0.36	0.22	1.66	14	0.12
School 8	0.05	0.11	0.41	14	0.69
School 9	0.11	0.10	1.11	14	0.28
School 10	-0.14	0.10	-1.47	14	0.16
School 11	0.04	0.09	0.39	14	0.70
School 12	0.38	0.20	1.91	14	0.08
School 13	0.24	0.20	1.22	14	0.24
School 14	-0.11	0.30	-0.37	14	0.71
Teacher with master’s degree	-0.04	0.05	-0.80	14	0.44
Teacher with prior experience with iPersonalize	-0.02	0.07	-0.24	14	0.82
Years of teaching experience	0.00	0.00	0.77	14	0.46
Student Level					
Female	0.08	0.02	3.73	1,088	< 0.001
Treatment by female interaction	-0.03	0.03	-1.11	1,088	0.27
Qualifying for free or reduced-price lunch	-0.02	0.02	-1.19	1,088	0.24
English learner	0.01	0.02	0.34	1,088	0.73
Qualifying for gifted services	-0.02	0.03	-0.45	1,088	0.65
Receiving special education services	-0.03	0.02	-1.25	1,088	0.21
Hispanic	0.04	0.03	1.56	1,088	0.12
Asian/Pacific Islander	-0.04	0.03	-1.60	1,088	0.11
Other race/ethnicity	0.07	0.04	1.59	1,088	0.11

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, female was uncentered, and all other predictors were centered on the classroom mean for the variable.

Table C8. HLM Coefficients for Model Predicting YouthTruth Score, Boys Only (*n* = 39 classrooms, 557 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	2.41	0.03	80.33	14	< 0.001
Treatment	0.06	0.05	1.31	14	0.21
Percentage of students qualifying for free or reduced-price lunch	0.31	0.17	1.85	14	0.09
Percentage of students identified as English learners	0.11	0.26	0.42	14	0.68
Percentage of students qualifying for gifted services	-0.04	0.10	-0.35	14	0.73
Percentage of students receiving special education services	-0.26	0.28	-0.93	14	0.37
Percentage Hispanic	0.20	0.26	0.77	14	0.46
Percentage Asian/Pacific Islander	0.57	0.36	1.57	14	0.14
Percentage Other race/ethnicity mean	0.04	0.45	0.10	14	0.93
School 1	0.30	0.15	2.00	14	0.07
School 2	0.29	0.17	1.65	14	0.12
School 3	0.04	0.12	0.35	14	0.73
School 4	0.11	0.16	0.69	14	0.51
School 5	0.46	0.15	2.98	14	0.01
School 6	0.19	0.13	1.51	14	0.15
School 7	0.22	0.16	1.34	14	0.20
School 8	0.12	0.13	0.97	14	0.35
School 9	0.10	0.13	0.83	14	0.42
School 10	-0.07	0.11	-0.63	14	0.54
School 11	0.06	0.10	0.59	14	0.56
School 12	0.53	0.22	2.42	14	0.03
School 13	0.07	0.16	0.44	14	0.67
Teacher with master's degree	-0.04	0.05	-0.79	14	0.44
Teacher with prior experience with iPersonalize	-0.02	0.07	-0.25	14	0.81
Years of teaching experience	0.00	0.00	0.55	14	0.59

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Student Level					
Qualifying for free or reduced-price lunch	-0.05	0.03	-1.42	511	0.16
English learner	-0.01	0.03	-0.27	511	0.79
Qualifying for gifted services	-0.03	0.05	-0.59	511	0.56
Receiving special education services	-0.06	0.03	-1.92	511	0.06
Hispanic	0.04	0.04	1.05	511	0.30
Asian/Pacific Islander	-0.06	0.04	-1.56	511	0.12
Other race/ethnicity	0.03	0.08	0.37	511	0.72

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C9. HLM Coefficients for Model Predicting YouthTruth Score, Girls Only (*n* = 40 classrooms, 580 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	2.52	0.03	79.54	14	< 0.001
Treatment	0.00	0.05	0.05	14	0.96
Percentage of students qualifying for free or reduced-price lunch	-0.04	0.24	-0.17	14	0.87
Percentage of students identified as English learners	0.06	0.17	0.34	14	0.74
Percentage of students qualifying for gifted services	0.02	0.11	0.18	14	0.86
Percentage of students receiving special education services	-0.03	0.37	-0.09	14	0.93
Percentage Hispanic	-0.39	0.29	-1.34	14	0.20
Percentage Asian/Pacific Islander	-0.39	0.25	-1.59	14	0.14
Percentage Other race/ethnicity mean	-0.59	0.57	-1.03	14	0.32
School 1	0.13	0.18	0.76	14	0.46
School 2	-0.04	0.25	-0.16	14	0.88
School 3	0.02	0.12	0.14	14	0.89
School 4	0.08	0.28	0.28	14	0.79

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
School 5	0.15	0.15	1.04	14	0.32
School 6	0.02	0.16	0.14	14	0.89
School 7	0.03	0.24	0.13	14	0.90
School 8	0.06	0.12	0.52	14	0.61
School 9	0.17	0.12	1.37	14	0.19
School 10	-0.17	0.14	-1.19	14	0.26
School 11	0.07	0.10	0.67	14	0.51
School 12	0.09	0.23	0.41	14	0.69
School 13	0.10	0.22	0.46	14	0.65
School 14	0.07	0.27	0.27	14	0.79
Teacher with master's degree	-0.03	0.06	-0.53	14	0.61
Teacher with prior experience with iPersonalize	-0.03	0.07	-0.43	14	0.67
Years of teaching experience	0.04	0.00	1.45	14	0.17
Student Level					
Qualifying for free or reduced-price lunch	-0.00	0.03	-0.18	533	0.86
English learner	0.03	0.03	0.93	533	0.35
Qualifying for gifted services	-0.00	0.04	-0.06	533	0.96
Receiving special education services	0.04	0.04	1.10	533	0.27
Hispanic	0.05	0.03	1.42	533	0.16
Asian/Pacific Islander	-0.01	0.03	-0.27	533	0.79
Other race/ethnicity	0.13	0.05	2.57	533	0.01

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Research Question 4—Is the Extent of Use of iPersonalize Associated with Increased Student Achievement?

Research question 4 was addressed with two sets of models. The first set of models included only students assigned to the treatment group and examined the association of the number of log-ons, the number of attempted quests, the number of completed quests, and the percentage of attempted quests that were completed with reading scale score (Tables C10–C13) and writing holistic score (Tables C14–C17).

Table C10. HLM Coefficients for Model Examining the Association Between the Number of Log-ons and Reading Scale Score (*n* = 21 classrooms, 601 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	594.02	1.08	550.24	2	< 0.001
Percentage of female students	12.70	61.21	0.21	2	0.86
Percentage of students qualifying for free or reduced-price lunch	39.91	74.06	0.54	2	0.64
Percentage of students identified as English learners	-17.89	50.61	-0.35	2	0.76
Percentage of students qualifying for gifted services	20.48	40.97	0.50	2	0.67
Percentage of students receiving special education services	72.67	64.20	1.13	2	0.38
Percentage Hispanic	-38.97	54.46	-0.72	2	0.55
Percentage Asian/Pacific Islander	-66.20	55.06	-1.20	2	0.35
School 1	35.95	24.50	1.47	2	0.28
School 3	-0.01	18.78	0.00	2	1.00
School 4	63.04	46.53	1.36	2	0.31
School 5	20.08	15.86	1.27	2	0.33
School 6	6.73	16.84	0.40	2	0.73
School 7	48.75	36.71	1.33	2	0.32
School 8	14.49	18.09	0.80	2	0.51
School 9	14.61	7.80	1.87	2	0.20
School 11	7.50	17.01	0.44	2	0.70
School 13	53.11	40.20	1.32	2	0.32
Pretest reading scale score classroom mean	0.79	0.36	2.16	2	0.16
Student Level					
Number of log-ons	-0.01	0.01	-1.15	571	0.25
Female	1.63	1.73	0.94	571	0.35
Qualifying for free or reduced-price lunch	-5.55	2.40	-2.31	571	0.02
English learner	-9.35	2.60	-3.60	571	< 0.001
Qualifies for gifted services	2.43	3.62	0.67	571	0.50
Receiving special education services	-4.36	2.82	-1.55	571	0.12

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Hispanic	-5.34	2.91	-1.83	571	0.07
Asian/Pacific Islander	1.22	2.96	0.41	571	0.68
Pretest reading scale score	0.67	0.02	28.65	571	< 0.001

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C11. HLM Coefficients for Model Examining the Association Between the Number of Attempted Quests and Reading Scale Score (*n* = 21 classrooms, 601 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	594.05	1.09	545.16	2	< 0.001
Percentage of female students	10.34	61.78	0.17	2	0.88
Percentage of students qualifying for free or reduced-price lunch	44.55	74.85	0.60	2	0.61
Percentage of students identified as English learners	-19.25	51.05	-0.38	2	0.74
Percentage of students qualifying for gifted services	17.85	41.39	0.43	2	0.71
Percentage of students receiving special education services	71.27	64.80	1.10	2	0.39
Percentage Hispanic	-41.71	54.97	-0.76	2	0.53
Percentage Asian/Pacific Islander	-63.32	55.62	-1.14	2	0.37
School 1	35.18	24.69	1.43	2	0.29
School 3	1.22	18.99	0.06	2	0.96
School 4	61.45	46.93	1.31	2	0.32
School 5	20.49	15.98	1.28	2	0.33
School 6	7.81	17.03	0.46	2	0.69
School 7	47.52	37.01	1.28	2	0.33
School 8	15.80	18.29	0.86	2	0.48
School 9	14.70	7.87	1.87	2	0.20
School 11	8.34	17.17	0.49	2	0.68
School 13	51.67	40.55	1.27	2	0.33
Pretest reading scale score classroom mean	0.82	0.37	2.22	2	0.16

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Student Level					
Number of attempted quests	-0.11	0.19	-0.61	570	0.54
Female	1.46	1.73	0.85	570	0.40
Qualifying for free or reduced-price lunch	-5.74	2.39	-2.40	570	0.02
English learner	-9.33	2.60	-3.59	570	< 0.001
Qualifying for gifted services	1.61	3.66	0.44	570	0.66
Receiving special education services	-4.30	2.81	-1.53	570	0.13
Hispanic	-5.46	2.91	-1.87	570	0.06
Asian/Pacific Islander	1.14	2.95	0.39	570	0.70
Pretest reading scale score	0.67	0.02	28.66	570	< 0.001

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C12. HLM Coefficients for Model Examining the Association Between the Number of Completed Quests and Reading Scale Score (n = 21 classrooms, 601 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	594.05	1.09	545.17	2	< 0.001
Percentage of female students	10.33	61.77	0.17	2	0.88
Percentage of students qualifying for free or reduced-price lunch	44.56	74.85	0.60	2	0.61
Percentage of students identified as English learners	-19.26	51.05	-0.38	2	0.74
Percentage of students qualifying for gifted services	17.84	41.39	0.43	2	0.71
Percentage of students receiving special education services	71.26	64.80	1.10	2	0.39
Percentage Hispanic	-41.72	54.96	-0.76	2	0.53
Percentage Asian/Pacific Islander	-63.32	55.61	-1.14	2	0.37
School 1	35.17	24.69	1.43	2	0.29
School 3	1.22	18.99	0.06	2	0.96
School 4	61.44	46.92	1.31	2	0.32
School 5	20.49	15.98	1.28	2	0.33

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
School 6	7.81	17.03	0.46	2	0.69
School 7	47.51	37.01	1.28	2	0.33
School 8	15.80	18.29	0.86	2	0.48
School 9	14.70	7.87	1.87	2	0.20
School 11	8.34	17.17	0.49	2	0.68
School 13	51.67	40.54	1.27	2	0.33
Pretest reading scale score classroom mean	0.82	0.37	2.22	2	0.16
Student Level					
Number of completed quests	0.23	0.18	1.31	570	0.19
Female	1.58	1.73	0.92	570	0.36
Qualifying for free or reduced-price lunch	-6.02	2.39	-2.52	570	0.01
English learner	-9.10	2.59	-3.51	570	< 0.001
Qualifying for gifted services	1.21	3.65	0.33	570	0.74
Receiving special education services	-4.10	2.81	-1.46	570	0.15
Hispanic	-4.99	2.92	-1.71	570	0.09
Asian/Pacific Islander	1.06	2.95	0.36	570	0.72
Pretest reading scale score	0.67	0.02	28.67	570	< 0.001

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C13. HLM Coefficients for Model Examining the Association Between the Percentage of Attempted Quests That Were Completed and Reading Scale Score (*n* = 21 classrooms, 601 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	594.05	1.09	545.22	2	< 0.001
Percentage of female students	10.25	61.75	0.17	2	0.88
Percentage of students qualifying for free or reduced-price lunch	44.65	74.84	0.60	2	0.61
Percentage of students identified as English learners	-19.34	51.02	-0.38	2	0.74

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Percentage of students qualifying for gifted services	17.79	41.37	0.43	2	0.71
Percentage of students receiving special education services	71.19	64.80	1.10	2	0.39
Percentage Hispanic	-41.83	54.93	-0.76	2	0.53
Percentage Asian/Pacific Islander	-63.32	55.61	-1.14	2	0.37
School 1	35.13	24.66	1.43	2	0.29
School 3	1.24	18.99	0.07	2	0.95
School 4	61.39	46.89	1.31	2	0.32
School 5	20.46	15.96	1.28	2	0.33
School 6	7.80	17.03	0.46	2	0.69
School 7	47.47	36.97	1.28	2	0.33
School 8	15.82	18.29	0.87	2	0.48
School 9	14.69	7.88	1.87	2	0.20
School 11	8.37	17.16	0.49	2	0.67
School 13	51.63	40.51	1.27	2	0.33
Pretest reading scale score classroom mean	0.82	0.37	2.22	2	0.16
Student Level					
Percentage of attempted quests that were completed	2.45	0.71	3.43	570	< 0.001
Female	1.80	1.71	1.05	570	0.29
Qualifying for free or reduced-price lunch	-5.80	2.37	-2.45	570	0.02
English learner	-9.54	2.57	-3.71	570	< 0.001
Qualifying for gifted services	1.62	3.61	0.45	570	0.66
Receiving special education services	-3.87	2.79	-1.39	570	0.17
Hispanic	-4.63	2.89	-1.61	570	0.11
Asian/Pacific Islander	1.29	2.92	0.44	570	0.66
Pretest reading scale score	0.66	0.02	27.96	570	< 0.001

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C14. HLM Coefficients for Model Examining the Association Between the Number of Logons and Writing Holistic Score (n = 21 classrooms, 569 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	3.85	0.08	49.87	2	< 0.001
Percentage of female students	13.18	17.72	0.74	2	0.54
Percentage of students qualifying for free or reduced-price lunch	-6.69	18.34	-0.37	2	0.75
Percentage of students identified as English learners	11.46	18.17	0.63	2	0.59
Percentage of students qualifying for gifted services	4.65	6.95	0.67	2	0.57
Percentage of students receiving special education services	-0.53	6.68	-0.08	2	0.94
Percentage Hispanic	8.97	15.03	0.60	2	0.61
Percentage Asian/Pacific Islander	2.97	3.37	0.88	2	0.47
School 1	3.96	3.43	1.15	2	0.37
School 3	1.42	1.46	0.98	2	0.43
School 4	3.31	4.11	0.81	2	0.51
School 5	4.49	3.64	1.23	2	0.34
School 6	2.64	2.30	1.15	2	0.37
School 7	4.87	4.47	1.09	2	0.39
School 8	-0.35	1.08	-0.33	2	0.77
School 9	1.25	1.23	1.02	2	0.42
School 11	-3.71	4.72	-0.79	2	0.51
School 13	2.52	3.52	0.72	2	0.55
Pretest writing holistic score classroom mean	-0.78	1.76	-0.44	2	0.70
Student Level					
Number of log-ons	0.00	0.00	-1.33	539	0.18
Female	-0.09	0.06	-1.51	539	0.13
Qualifying for free or reduced-price lunch	0.03	0.08	0.40	539	0.69
English learner	-0.32	0.09	-3.51	539	< 0.001
Qualifying for gifted services	-0.15	0.12	-1.23	539	0.22
Receiving special education services	-0.33	0.10	-3.23	539	< 0.01

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Hispanic	-0.12	0.10	-1.18	539	0.24
Asian/Pacific Islander	0.08	0.10	0.81	539	0.42
Pretest writing holistic score	0.47	0.04	11.33	539	< 0.001

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered around the classroom mean for the variable.

Table C15. HLM Coefficients for Model Examining the Association Between the Number of Attempted Quests and Writing Holistic Score (n = 21 classrooms, 569 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	3.85	0.08	49.35	2	< 0.001
Percentage of female students	12.99	17.90	0.73	2	0.54
Percentage of students qualifying for free or reduced-price lunch	-6.68	18.52	-0.36	2	0.75
Percentage of students identified as English learners	11.34	18.36	0.62	2	0.60
Percentage of students qualifying for gifted services	4.57	7.02	0.65	2	0.58
Percentage of students receiving special education services	-0.55	6.75	-0.08	2	0.94
Percentage Hispanic	8.91	15.18	0.59	2	0.62
Percentage Asian/Pacific Islander	2.98	3.40	0.88	2	0.47
School 1	3.90	3.47	1.13	2	0.38
School 3	1.41	1.47	0.96	2	0.44
School 4	3.22	4.16	0.77	2	0.52
School 5	4.42	3.68	1.20	2	0.35
School 6	2.60	2.32	1.12	2	0.38
School 7	4.77	4.51	1.06	2	0.40
School 8	-0.36	1.09	-0.33	2	0.77
School 9	1.24	1.24	1.00	2	0.42
School 11	-3.66	4.77	-0.77	2	0.52
School 13	2.44	3.56	0.69	2	0.56
Pretest writing holistic score classroom mean	-0.76	1.78	-0.43	2	0.71

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Student Level					
Number of attempted quests	0.00	0.01	0.42	538	0.67
Female	-0.09	0.06	-1.42	538	0.16
Qualifying for free or reduced-price lunch	0.02	0.08	0.24	538	0.81
English learner	-0.32	0.09	-3.43	538	< 0.001
Qualifying for gifted services	-0.16	0.13	-1.26	538	0.21
Receiving special education services	-0.33	0.10	-3.24	538	< 0.01
Hispanic	-0.12	0.10	-1.16	538	0.25
Asian/Pacific Islander	0.07	0.10	0.72	538	0.47
Pretest writing holistic score	0.46	0.04	11.08	538	< 0.001

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C16. HLM Coefficients for Model Examining the Association Between the Number of Completed Quests and Writing Holistic Score (*n* = 21 classrooms, 569 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	3.85	0.08	49.34	2	< 0.001
Percentage of female students	13.00	17.90	0.73	2	0.54
Percentage of students qualifying for free or reduced-price lunch	-6.68	18.52	-0.36	2	0.75
Percentage of students identified as English learners	11.34	18.36	0.62	2	0.60
Percentage of students qualifying for gifted services	4.57	7.02	0.65	2	0.58
Percentage of students receiving special education services	-0.55	6.75	-0.08	2	0.94
Percentage Hispanic	8.91	15.18	0.59	2	0.62
Percentage Asian/Pacific Islander	2.98	3.40	0.88	2	0.47
School 1	3.90	3.47	1.13	2	0.38
School 3	1.41	1.47	0.96	2	0.44
School 4	3.22	4.16	0.77	2	0.52
School 5	4.42	3.68	1.20	2	0.35

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
School 6	2.60	2.32	1.12	2	0.38
School 7	4.77	4.51	1.06	2	0.40
School 8	-0.36	1.09	-0.33	2	0.77
School 9	1.24	1.24	1.00	2	0.42
School 11	-3.66	4.77	-0.77	2	0.52
School 13	2.44	3.56	0.69	2	0.56
Pretest writing holistic score classroom mean	-0.76	1.78	-0.43	2	0.71
Student Level					
Number of completed quests	0.01	0.01	1.02	538	0.31
Female	-0.09	0.06	-1.40	538	0.16
Qualifying for free or reduced-price lunch	0.02	0.08	0.21	538	0.83
English learner	-0.32	0.09	-3.44	538	< 0.001
Qualifying for gifted services	-0.16	0.13	-1.29	538	0.20
Receiving special education services	-0.33	0.10	-3.23	538	< 0.01
Hispanic	-0.11	0.10	-1.09	538	0.28
Asian/Pacific Islander	0.08	0.10	0.73	538	0.46
Pretest writing holistic score	0.46	0.04	10.95	538	< 0.001

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

Table C17. HLM Coefficients for Model Examining the Association Between the Percentage of Attempted Quests That Were Completed and Writing Holistic Score (*n* = 21 classrooms, 569 students)

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Classroom Level					
Intercept	3.85	0.08	49.34	2	< 0.001
Percentage of female students	13.00	17.90	0.73	2	0.54
Percentage of students qualifying for free or reduced-price lunch	-6.68	18.53	-0.36	2	0.75
Percentage of students identified as English learners	11.34	18.36	0.62	2	0.60

Predictor	Coefficient	Standard Error	*t* Ratio	Approx. *df*	*p* Value
Percentage of students qualifying for gifted services	4.57	7.02	0.65	2	0.58
Percentage of students receiving special education services	-0.55	6.75	-0.08	2	0.94
Percentage Hispanic	8.91	15.18	0.59	2	0.62
Percentage Asian/Pacific Islander	2.98	3.40	0.88	2	0.47
School 1	3.90	3.47	1.13	2	0.38
School 3	1.41	1.47	0.96	2	0.44
School 4	3.22	4.16	0.77	2	0.52
School 5	4.42	3.68	1.20	2	0.35
School 6	2.60	2.32	1.12	2	0.38
School 7	4.77	4.51	1.06	2	0.40
School 8	-0.36	1.09	-0.33	2	0.77
School 9	1.24	1.24	1.00	2	0.42
School 11	-3.66	4.77	-0.77	2	0.52
School 13	2.44	3.56	0.69	2	0.56
Pretest writing holistic score classroom mean	-0.76	1.78	-0.43	2	0.71
Student Level					
Female	-0.09	0.06	-1.38	538	0.17
Percentage of attempted quests that were completed	0.04	0.03	1.58	538	0.11
Qualifying for free or reduced-price lunch	0.02	0.08	0.26	538	0.80
English learner	-0.32	0.09	-3.53	538	< 0.001
Qualifying for gifted services	-0.16	0.13	-1.24	538	0.22
Receiving special education services	-0.33	0.10	-3.22	538	< 0.01
Hispanic	-0.11	0.10	-1.05	538	0.29
Asian/Pacific Islander	0.08	0.10	0.79	538	0.43
Pretest writing holistic score	0.46	0.04	10.99	538	<0.001

Note. At the classroom level, treatment was uncentered, and all other predictors were centered on the mean for the whole sample. At the student level, all predictors were centered on the classroom mean for the variable.

The second set of models included data from students from both the treatment and control groups. Two modeling approaches were used: instrumental variables and continuous ASPES (CASPES). Table C18 displays the results of instrumental variable models estimated with 2SLS models (with randomization indicator as the instrument) and CASPES models.

The significance values for the 2SLS models were based on clustered standard errors, with clustering at the classroom level. Usage in the control group was coded to 0 for these analyses. The CASPES models were based on HLM with clustering at the classroom level and the predicted usage values, an indicator for treatment group membership, and the interaction between treatment and predicted usage as the independent variables.

The 2SLS results presented in Table C18 include only the pretests as additional exogenous variables. The CASPES results use all available covariates *except* for the pretests in the first-stage models that created predicted usage variables. Pretests were used only in the second-stage CASPES models that estimated variation in treatment effects as a function of usage. This approach was taken to minimize collinearity between the predicted usage variable and other variables in the second-stage regression. The usage variable (or predicted usage variable in the case of CASPES) was converted to a *z* score before models were fit so that the effect sizes reported can be interpreted as the effect sizes associated with a standard deviation change in usage.

When reading was the dependent variable, the instrumental variable results showed small but statistically significant effects (Table C18). Effects were of a similar size, regardless of how usage was defined. However, when all exogenous variables were included in both regressions, effect sizes were reduced to virtually zero in all cases, with *p* values greater than .90 (not shown). The CASPES analyses showed virtually zero influence of usage on treatment effects (Table C18). Roughly zero estimated effects were also found when all covariates were added to both the first- and second-stage regressions (not shown). Overall, the evidence pointed to treatment effects close to zero, regardless of the level of usage.

When writing was the dependent variable, the instrumental variable analyses showed somewhat larger estimated positive effects of usage (Table C18). However, none was statistically significant. Additionally, when the other exogenous variables were added to the 2SLS regressions, estimated treatment effects were negative and not statistically significant (not shown). The CASPES analyses were all consistent with zero average impact of treatment, regardless of usage level.

Table C18. Summary of Results of Models Examining the Causal Relationship Between Student Usage Variables and Academic Outcomes

Usage Variable	Instrumental Variables		CASPES	
	Effect Size	Significance	Effect Size	Significance
Dependent Variable: Reading Scale Score				
Number of log-ons	0.086	0.038	-0.050	0.236
Number of quests attempted	0.050	0.019	-0.006	0.870
Number of quests completed	0.053	0.019	-0.011	0.791

Usage Variable	Instrumental Variables		CASPES	
	Effect Size	Significance	Effect Size	Significance
Percentage of attempted quests that were completed	0.047	0.021	0.035	0.414
Dependent Variable: Holistic Writing Score				
Number of log-ons	0.184	0.106	-0.012	0.910
Number of quests attempted	0.107	0.105	-0.070	0.500
Number of quests completed	0.114	0.103	0.050	0.622
Percentage of attempted quests that were completed	0.101	0.108	0.116	0.281

Note. Effect size was computed as the estimated regression coefficient divided by the unconditional standard deviation of the dependent variable (49.30 in the case of reading, and 1.02 in the case of writing). In CASPES models, the effect size represents the estimated difference in the standardized average treatment effects between students predicted to use at rate u and students predicted to use at rate $u + 1$. Usage variables are in z score units.

Research Question 5—What Is the Impact of iPersonalize on Low and High Quantiles of Sixth-Grade Student Achievement in ELA?

Research question 6 was addressed with a series of quantile regressions. The results of these analyses are summarized in Tables C19 and C20.

Table C19. Summary of Quantile Mixed Models Examining Impact of Randomization to iPersonalize on Academic Outcomes at Percentiles of Conditional Achievement Distribution

Quantile	Reading Outcome			Writing Outcome		
	Effect Size	Lower Limit	Upper Limit	Effect Size	Lower Limit	Upper Limit
0.1	0.055	-0.034	0.144	0.031	-0.326	0.389
0.2	0.055	-0.034	0.144	-0.002	-0.360	0.355
0.3	0.055	-0.034	0.144	-0.049	-0.405	0.308
0.4	0.055	-0.034	0.144	-0.050	-0.445	0.346
0.5	0.055	-0.034	0.144	0.110	-0.265	0.484
0.6	0.055	-0.034	0.144	0.127	-0.209	0.463
0.7	0.055	-0.034	0.144	0.075	-0.230	0.382
0.8	0.055	-0.034	0.144	0.189	-0.082	0.459

Quantile	Reading Outcome			Writing Outcome		
	Effect Size	Lower Limit	Upper Limit	Effect Size	Lower Limit	Upper Limit
0.9	0.055	-0.034	0.144	0.196	-0.056	0.449

Note. Effect size was computed as the estimated regression coefficient from the quantile mixed model divided by the unconditional standard deviation of the dependent variable (49.30 in the case of reading, and 1.02 in the case of writing). Confidence intervals did not account for sampling variation in the estimated standard deviation. Lower and upper 95 percent confidence interval limits are reported.

Table C20. Summary of Quantile Mixed Models Examining Impact of Randomization to iPersonalize on Engagement at Percentiles of Conditional Engagement Distribution.

Quantile	YouthTruth Score		
	Effect Size	Lower Limit	Upper Limit
0.1	-0.516	-2.008	0.976
0.2	0.190	-0.200	0.580
0.3	0.170	-0.140	0.472
0.4	0.152	-0.144	0.448
0.5	0.188	-0.104	0.484
0.6	0.188	-0.072	0.444
0.7	0.216	-0.028	0.460
0.8	0.214	-0.096	0.340
0.9	2.536	0.564	4.520

Note. Effect size was computed as the estimated regression coefficient from the quantile mixed model divided by the unconditional standard deviation of the dependent variable (0.25). Confidence intervals did not account for sampling variation in the estimated standard deviation. Lower and upper 95 percent confidence interval limits are reported.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Author.

Angrist, J. D., and Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*(4), 69–85. doi:10.1257/jep.15.4.69

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427–445. doi:10.1016/j.jsp.2006.04.002

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. doi:10.18637/jss.v067.i01

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a taxonomy linking game attributes to learning: An empirical study. *Simulation and Gaming*, *43*(6), 729–760. doi:10.1177/1046878112439444

Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*. New York, NY: William T. Grant Foundation. Retrieved from https://eric.ed.gov/?id=ED568396

Curriculum Associates. (n.d.) *The science behind i-Ready's adaptive diagnostic*. North Billerica, MA: Author. Retrieved from http://www.casamples.com/downloads/i-Ready-Diagnostic-Position-Paper.pdf

Curriculum Associates. (2017). *What growth should I expect my students to make in i-Ready? (2017–2018 school year)*. North Billerica, MA: Author. Retrieved from https://www.sjsd.k12.mo.us/cms/lib/MO01001773/Centricity/Domain/63/iReady%20Expected%20Growth%20in%20Scale%20Scores.pdf

Educause (2009). *7 things you should know about alternate reality games*. Retrieved from https://library.educause.edu/resources/2009/1/7-things-you-should-know-about-alternate-reality-games

Educational Research Institute of America. (2014). *I-Ready diagnostic New York State validity study*. Bloomington, IN: Author. Retrieved from http://www.scotiaglenvilleschools.org/emailforms/i_ReadyNY_validitystudy.pdf

Fearrington, J. Y., Parker, P. D., Kidder-Ashley, P., Gagnon, S. G., McCane-Bowling, S., & Sorrell, C. A. (2014). Gender differences in written expression curriculum-based

measurement in third- through eighth-grade students. *Psychology in the Schools*, *51*(1), 85–96. doi:10.1002/pits.21733

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. doi:10.3102/00346543074001059

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In S. L. Christensen, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). New York, NY: Springer Science+Business Media.

Gennetian, L. A., Morris, P. A., Bos, J. M., & Bloom, H. S. (2005). Constructing instrumental variables from experimental data to explore how treatments produce effects. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 75–114). New York, NY: Russell Sage.

Geraci, M. (2014). Linear quantile mixed models: The lqmm package for Laplace quantile regression. *Journal of Statistical Software, 57*(13), 1–29. doi:10.18637/jss.v057.i13

Geraci, M., & Bottai, M. (2014). Linear quantile mixed models. *Statistics and Computing, 24*(3), 461–479. doi:10.1007/s11222-013-9381-9

Institute of Education Sciences. (2018). *Request for applications: Low-cost, short-duration evaluation of education interventions* (CDFA No. 84.305L). Washington, DC: U.S. Department of Education. Retrieved from https://ies.ed.gov/funding/pdf/2019_84305L.pdf

Institute of Education Sciences. (2019). *Request for applications: Education research grants* (CFDA No. 84.305A). Washington, DC: Department of Education. Retrieved from https://ies.ed.gov/funding/pdf/2020_84305A.pdf

Jiang, Y., & Small, D. (2015). *Package 'ivpack'* (Reference manual). Retrieved from https://cran.r-project.org/web/packages/ivpack/ivpack.pdf

Kleiber, C., & Zeileis, A. (2008). *Applied econometrics with R*. New York, NY: Springer Science+Business Media.

Kuznetsova, A., Brockhoff, P. R., & Christensen, R. (2019). *Package 'lmerTest'* (Reference manual). Retrieved from https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf

Landers, R. N. (2015). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming*, *45*(6), 752–768. doi:10.1177/1046878114563660

Lenhart, A., Kahne, J., Middaugh, E., Macgill, A. R., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics: Teens' gaming experiences are diverse and include significant social interaction and civic engagement.* Washington, DC: Pew Internet and American Life Project. Retrieved from https://eric.ed.gov/?id=ED525058

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. Cambridge, MA: MIT Press.

Moulton, S. R., Peck, L. R., & Greeney, A. (2018). Analyzing the influence of dosage in social experiments with application to the supporting healthy marriage program. *American Journal of Evaluation, 39*(2), 257–277. doi:10.1177/1098214017698566

National Center for Education Statistics. (2013). *The Nation's Report Card: Trends in academic progress 2012* (NCES 2013-456). Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2013456

Peck, L. R. (2003). Subgroup analysis in social experiments: Measuring program impacts based on post-treatment choice. *American Journal of Evaluation, 24*(2), 157–187. doi:10.1016/S1098-2140(03)00031-6

Peck, L. R. (2015). Using impact evaluation tools to unpack the black box and learn what works. *Journal of MultiDisciplinary Evaluation, 11*(24), 54–67. Retrieved from http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/415

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Vantage Learning. (2006). *Research summary: IntelliMetric scoring accuracy across genres and grade levels*. Newton, PA: Author. Retrieved from http://www.vantagelearning.com/docs/intellimetric/IM_ReseachSummary_IntelliMetric_Accuracy_Across_Genre_and_Grade_Levels.pdf

What Works Clearinghouse. (2017). *Procedures handbook* (Version 4.0). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from https://ies.ed.gov/ncee/wwc/handbooks

YouthTruth. (n.d.). *YouthTruth student survey: Design and methodology*. San Francisco, CA: Author. Retrieved from http://www.youthtruthsurvey.org/wp-content/uploads/2016/06/YouthTruth-Design-and-Methodology-Report-2016.pdf

The research department at Marzano Research supports partners in improving education systems, practices, and outcomes for all learners.

Founded in 2008, Marzano Research began working with state and local education organizations and practitioners to understand the challenges they face and support them in defining the questions, conducting the research, and implementing the answers to enhance educational results.

Today, Marzano Research has grown to become one of the leading research organizations in the country, providing rigorous research, evaluation, and technical assistance to federal, state, local, and private partners. As part of that work, we serve as the lead for the Regional Education Laboratory in the central region, working with state and local education agencies in seven states as thought partners and researchers to address some of the most challenging issues in education.

Marzano Research
12577 E Caley Ave
Centennial, CO 80111
research@marzanoresearch.com
720.463.3600